THE UNIVERSITY OF CHICAGO

ESSAYS ON SOCIAL INFLUENCES IN PUBLIC ECONOMICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
JUSTIN ERIC HOLZ

CHICAGO, ILLINOIS
JUNE 2021

To Nicole

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my committee, Christopher Blattman, Leonardo Bursztyn, Steven Durlauf, and John List for their exceptional mentorship. Many other professors at University of Chicago and elsewhere provided help that, in one way or another, was instrumental for the execution of this dissertation. I thank Julio Elías, Yana Gallen, Brent Hickman, Min Lee, Steven Levitt, Michael Price, and Andrew Zentner.

Many others have been integral to these and other projects as well as my overall well-being throughout graduate school. I am especially grateful to Dan Alexander, Bocar Ba, Alec Brandon, Luigi Butera, Cynthia Cook-Conley, Michael Cuna, Jonathan Davis, John Donmoyer, Mariella Gonzales, Clark Halliday, Erik Johnson, Chien-yu Lai, Seunghoon Lee, Alexandr Lenk, Ariel Listo, Claire Mackevicius, Maria Adelaida Martinez, Val Michelman, Fatemeh Momeni, Miguel Morales, David Novgorodsky, Joe Seidel, Andrew "Rusty" Simon, Diana Smith, Dana Suskind, Jamie Temmer, Mattie Toma, Haruka Uchida, Wendy Wong, Derek Wu, and Karen Ye. John Cadigan, Jean Fletcher, Timothy Opiela, Drew Murphy, and Charles Weise were instrumental in setting me off on the right academic footing.

Finally, I am indebted to my wife, Nicole Ozminkowski, and my mother, Susan Holz. This dissertation would not have been possible without their incredible support.

ABSTRACT

This dissertation studies the role of peer effects, social preferences, and social shaming in decision-making. The first chapter, co-authored with Bocar Ba and Roman Rivera, studies peer effects in a police officer's decision to use force. This chapter shows that peer injuries increase the probability of using force by 7% in the subsequent week. Officers are also more likely to injure suspects and receive complaints about neglecting victims and violating suspects' constitutional rights. The effect is concentrated in a narrow time window near the event and is not associated with significantly lower injury risk to the officer. Together, these findings suggest that emotional responses by the officers might drive the increase in force.

The second chapter, co-authored with Rafael Jiménez-Durán and Eduardo Laguna-Müggenburg, demonstrates how social preferences can lead consumers to enforce price-ceilings. Thirty-four states have laws against increasing the price of necessary goods during crises, and individuals take costly actions to report violators of these laws. We provide a theoretical explanation for how these consumer reports reflect a willingness to pay to prevent third-party transactions, called repugnance by Roth (2007). We then introduce a new experimental paradigm to measure revealed preferences against illicit transactions with imperfect enforcement, called an Incentivized Reporting Experiment (IRE). We implement IRE during the first wave of COVID-19 to measure the willingness to pay to report sellers who increase personal protective equipment prices. Subjects in the experiment pay to report sellers for price-gouging even when they would be willing to buy the goods at the sellers' price. Subjects also support third-party transactions at high prices even when they are willing to pay to report the seller. The mechanism driving the repugnance towards price-gouging is good specific. Individuals oppose the price-gouging of face masks out of a concern for external consumption. In contrast, individuals oppose the price-gouging of hand sanitizer due to a distaste for firm profits.

The final chapter, co-authored with Alejandro Zentner, John List, Marvin Cardoza, and

Joaquin Zentner, explores the effect of social shaming on tax compliance in the Dominican Republic. We use two large-scale natural field experiments to examine the effectiveness of deterrence nudges on tax compliance in the Dominican Republic. In collaboration with the tax authority, we sent messages to 84,500 businesses who collectively paid \$800 million in the year before the experiment. We provide the first experimental evidence that social shaming is an effective way to increase tax compliance. However, it is much less effective than a prison deterrence message. Overall, we find that increasing the salience of prison sentences or the public disclosure of evasion increases tax revenue by \$193 million (0.23% of GDP). Using a unique sample of large firms, we show that the largest firms, who pay 84% of all corporate income taxes, are considerably more responsive to nudges than typically-studied smaller firms.

CHAPTER 1

PEER EFFECTS IN POLICE USE OF FORCE

with Roman G. Rivera and Bocar A. Ba[1]

Abstract

We study the link between officer injuries on duty and their peers' propensity to employ force using a network of officers who attended the police academy together through a random lottery. Peer injuries on duty increase the probability of using force by 7% in the subsequent week. Officers are also more likely to injure suspects and receive complaints about neglecting victims and violating suspects' constitutional rights. The effect is concentrated in a narrow time window near the event and is not associated with significantly lower injury risk to the officer. Together, these findings suggest that emotional responses by the officers might drive the increase in force.

1. We thank Amani Abou Harb, Peter Arcidiacono, Pat Bayer, Sandra Black, Chris Blattman, Alec Brandon, Leonardo Bursztyn, Cynthia Cook Conley, Jennifer Doleac, Robert Dur, Steven Durlauf, Yana Gallen, Robert Garlick, Mariella Gonzalez, Kareem Haggag, Robert Lalonde, Mariana Laverde, John List, Miguel Morales, David Novgorodsky, Emily Owens, Nicole Ozminkowski, Andrew Papachristos, Devin Pope, Wendy Wong, Karen Ye as well as seminar participants at the University of Chicago, the Texas Economics of Crime Workshop, The Association for Public Policy Analysis & Management and the Conference on Empirical Legal Studies. We thank Sam Stecklow, the Invisible Institute, and Craig Futterman for help with the data. We thank UC Irvine and Duke Economics Departments for generous financial support. Our study was approved by the University of Chicago Institutional Review Board, IRB21-0524.

1.1 Introduction

Police misconduct and use of force is a critical social and economic issue in the United States. Over one thousand people are killed by police each year, far more than any other industrial democracy (Lartey, 2015). That means that for every ten murders, one civilian dies during a police encounter (Annan-Phan and Ba, 2020). The rate of police killings is exceptionally high among African Americans who face lifetime odds of being killed by the police up to one in a thousand (Edwards et al., 2019) and experience reductions in human capital attainment from police killings (Ang, 2021). Moreover, police misconduct, including wrongful death, resulted in over $3 billion in lawsuits over the past ten years (Thomson-DeVeaux et al., 2021).

The social and economic damages caused by police activity has led the Biden Administration to view addressing "systemic police misconduct" as a key policy objective (*Joe Biden*, 2020). However, enacting policies that reduce police killings and misconduct requires first understanding the factors that cause an officer to use force. This paper advances this discussion by showing peer effects are an important mechanism driving an officer's decision to use force. Specifically, we use highly detailed administrative data from the Chicago Police department to show that officer injuries increase their peer's propensity to use force.

Identifying the effect of an officer injury on a peer's decision requires exogeneity in the officer's injury with an assignment probability independently distributed across groups (Angrist, 2014). We approximate this ideal by combining random assignment of officers to groups with a difference-in-differences design exploiting exogenous timing of officer injuries. In our setting, the Chicago Police Department randomly assigns new officers to police academy cohorts using a random lottery. After the academy is over, these officers move to different districts across the city. This process creates a group of officers who attended the police academy together but now work in a different part of the city, which we call former peers.[2]

2. Shue (2013) and Ager et al. (2018) use a similar definition to define their peer groups.

We then estimate the effect of officer injuries by comparing the change in an officer's outcomes over time between officers who did and did not have a former peer injured in the previous week. Throughout the analysis, we use outcomes without conditioning on whether a suspect interacts with the police to avoid the issues that arise when conditioning on intermediate outcomes (Heckman and Durlauf, 2020; Fryer, 2020).

We find that on-duty injuries increase a former peer's propensity to use force by seven percent in the subsequent week, leading to more civilian injuries. Officers are also more likely to receive a complaint about a false arrest or improper search in the week following a peer injury. This finding suggests that officers respond to the injury of a former peer by violating suspects' constitutional rights. We also find a higher probability of receiving a complaint about the failure to provide service, suggesting officers sort out of helping potential victims.

These results likely understate the effect of officer injuries for two reasons. First, the estimation strategy excludes peers in the same geographic police district because these officers may experience correlated shocks to civilian non-compliance. Injuries to these peers are likely at least as impactful due to plausibly lower social distance between the officers. Indeed, when we consider another group with plausibly lower social distance – academy members of the same race – our effects are twice double in magnitude. Second, we drop the first year of data for all officers in our sample because we cannot identify the officer's geographic district during that time. Effects are likely larger during this period because officers have less experience and are likely closer to their academy peers.

To better understand these effects, we document several moderating effects of peer injuries. First, the effects are driven primarily by Black and male suspects. Second, as in Guryan et al. (2009), we find that professional experience attenuates social influences such as peer effects. More experienced officers are less responsive to former peers' injuries in terms of using force, injuring suspects, or acting in a manner that causes a civilian to issue a complaint.

Next, we explore why officers respond to peer injuries. Similar to Card and Dahl (2011) and Munyo and Rossi (2015), the effects on police violence are immediate and quickly disappear two weeks after the event. The immediate effects that quickly fade in conjunction with the moderating effects of tenure suggest that emotional responses are driving the results. This result is similar to Ta et al. (2021) who find that a police officer's emotional reactivity is lower in more experienced police officers. We can rule out several alternatives. First, we use force-use unaccompanied by a peer injury to show that these results cannot be explained by officers mimicking the use-of-force behavior of their peers. Next, we link officers to arrest records and rule out officers responding to peer injuries by changing their effort using a similar measure as Mas (2006) and Ba and Rivera (2019). Finally, we find that officers are no less likely to be injured in the week after a former peer injury. This finding is inconsistent with officers using information about their peer's injury to update beliefs about the risk of on-duty injuries.

This article makes three contributions. First, it documents that peer injuries increase officers' propensity to use force, injure suspects, and violate suspect's constitutional rights. This finding contributes to the vast literature that focuses on the police's effect on crime, remaining agnostic about the police officers' production function.[3] More recent studies have documented how aggressive policing can reduce the educational performance of minority groups, negatively affect attitudes toward the state and undermine police legitimacy.[4] By identifying a causal determinant in the decision to use force, this paper builds upon the burgeoning literature attempting to unpack the black box of police productivity by providing evidence that police officers respond to their peers' outcomes.[5] These findings introduce a

3. See Levitt (2002), Di Tella and Schargrodsky (2004), Evans and Owens (2007), Draca et al. (2011), Chalfin and McCrary (2018), and Morales (2020) for papers documenting the effect of policing on crime.

4. See skolnick1993above, Tyler (2004), Weitzer and Tuch (2004), Brunson and Miller (2005), Lum and Nagin (2017), Manski and Nagin (2017), Legewie and Fagan (2019), and Ang (2021).

5. See Fryer (2018), Owens et al. (2018), Ba and Rivera (2019), Ba et al. (2020), Annan-Phan and Ba (2020), and Zimring (2019) for other work attempting to uncover the determinants of police force.

new dimension for policymakers to consider. Policies that increase the risk of injury to officers will have a muted effect on force use when peers respond by increasing force. Alternatively, policies that reduce the risk to officers will have the added benefit of decreasing the amount of force used by officers.

Second, it contributes to the literature on peer effects by documenting evidence that individuals respond to peers' outcomes rather than choices in the workplace (Mas and Moretti, 2009; Cornelissen et al., 2017). This result suggests that direct responses to peer outcomes may partially drive the results in other studies that find negative spillovers. For example, Carrell and Hoekstra (2010) find negative spillovers from children in troubled families and argue that these effects operate through the reduced achievement or increased disruption of the affected child. Similarly, Murphy (2019) attributes contemporaneous misconduct in the military to peers responding to the poor behavior of other soldiers. Our finding provides new mechanisms for exploring such results. We also confirm that one's peer group can affect individuals' choices and outcomes long after the group dissipates (Bayer et al., 2009; Shue, 2013) and that professional experience attenuates social influences (Guryan et al., 2009).

Third, this paper contributes to the literature on the effects of exposure to violence by showing that a peer's exposure to violence affects an individual's behavior in high-stakes decisions. Lab and artefactual field experiments have uncovered evidence that exposure to violence can increase preferences for certainty and impatience while decreasing emotional regulation.[6] Outside of the lab, Bauer et al. (2016) find that people exposed to war behave more cooperatively and altruistically towards their ingroup. In contrast, we find that those exposed to violence behave less altruistically towards outgroup members. Similarly, Hjort (2014) find that ethnic conflict increases animus against outgroup members.

The remainder of the paper proceeds as follows. Section 1.2 describes the institutional

6. See Callen et al. (2014), Imas et al. (2018), Moya (2018), Brown et al. (2019) for lab and artefactual field evidence that violence can affect preferences and behavior. Osofsky (1997) finds that violence can decrease emotional regulation.

background, providing information on the network formation and the policy governing use-of-force decisions. Section 1.3 describes the relevant datasets, sample definitions, and summary statistics. Section 1.4 explains the research design used to generate the estimates provided in section 1.5. Section 1.6 sheds light on the mechanisms suggested by auxiliary data analysis. Section 1.7 concludes.

1.2 Background

1.2.1 Formation of Police Networks

The Chicago Police Department's recruitment process creates an ideal setting to study spillover effects. The recruitment process generally follows five steps: (1) a recruitment call, (2) an entrance exam, (3) a referral lottery, (4) a battery of physical and mental tests, and finally (5) the officer attends the police academy.[7]

The CPD regularly issues recruitment calls. Table 1.1 displays the nine recruitment calls made between 2002 and 2013. After applying, prospective officers take an exam meant to evaluate the officer's cognitive and non-cognitive abilities. In Step 3, the CPD adds all of the applicants who pass the exam to an eligibility list and provides each a lottery number. These applicants are referred to the CPD academy in lottery order as vacancies become available. Applicants remain on the lottery list until it is either exhausted or retired (Chicago Police Department, 2016), with veterans receiving priority in the randomization. This application process ensures that individuals did not select into specific cohorts based on the propensity to use force, be injured, or respond to peer injuries with violence.

Our data contain officers' start dates and the dates that the CPD held recruitment tests. However, we do not observe an individual officer's test date as the CPD is unable to supply

7. Officers who apply to be part of the Chicago Police Department (CPD) must fulfill age and citizenship requirements. Applicants must also have a combination of post-secondary and/or army training: applicants must have least sixty semester hours from an accredited university, three years of active duty in the armed forces or thirty semester hours, and one continuous year of active duty.

Table 1.1: Police Entrance Lotteries

Exam	Dates of Administration	Attended	Passed	Classes	Officers
2002	1/12/2002	3,150	No info	16	268
2003	11/22/2003	No	No info	5	28
2004	11/20/2004	4,163	No info	7	317
2005	2/18/2006; 2/19/2006	4,061	3,338	4	174
2006-1	6/4/2006	1,508	1,255	2	139
2006-2	8/6/2006	1,025	863	3	181
2006-3	11/5/2006	1,795	1,487	18	810
2010	12/11/2010 makeups: 3/12/2011; 6/11/2011; 9/25/2011; 12/3/2011; 6/2/2013; 12/1/2012; 3/9/2013	8,621	7,689	26	1237
2013	12/14/2013 military makeups 6/28/2014; 12/7/2014; 6/13/2015; 12/6/2015	14,788	12,877	2	141

Note: Sample includes officers starting the police academy between January 2002 and December 2013.

this information (see Section 1.8.7 for more information). In its place, we use the date of the most recent test before the officer started working as a proxy for their test date. We overcome any measurement error in our proxy test dates by using individual fixed effects in our main specifications. These individual fixed effects allow us to overcome this issue because whatever date the officer took the test is time-invariant and subsumed by the individual fixed effect.

Applicants whose numbers are called proceed to Step 4, where they must pass further examinations to proceed. These include a physical test, a background check, a psychological evaluation, and a drug test. After the officer passes these examinations, they start at the police academy. We refer to all officers starting the academy together in the same month as an academy cohort. The cohorts start at the police academy throughout the sample period. On average, police academy cohorts are 78% male, 49% white, 17% black, and 34% Hispanic. The median age of new officers in our sample is 28. Cohorts have on average 42.81 individuals, though sizes range considerably.

Once applicants enter the academy, the Education and Training Division provides over 900 hours of basic training over six months. Training includes instruction on the use of force tactics, including firearms and control techniques. There is also physical and scenario-based training in the classroom. CPD recruits receive extra training on gangs, drugs, law, ethics, report writing, vehicle stops, use of force, and driving.

After completing the academy, officers complete roughly twelve months of probationary field training.[8] During the twelve months of field training, the CPD assigns probationary officers to districts at their discretion. Duty assignments can change day-to-day during this period. Unfortunately, the unit assignment data does not record probationary assignments. After the probationary period is over, officers move to more permanent police units, based on the needs of the CPD rather than the preferences of the police officer. We define an officer's peer group as their "former peers", meaning the members of their academy cohort who are working in a different geographic unit in a given week.

Most police officers work in geographic units which are tasked with policing a specific geographic area, known as a police district. In 2013, there were 22 police districts in Chicago, with officers often being assigned to police specific beats (about one square mile) within the district.[9] We focus our analysis on these geographic units. In an average week, we observe 94 officers who joined between 2002 and 2013 in each of the 22 units.

1.2.2 Use of Force Policy

The Chicago Police Department defines force as physical contact by a Department member used to compel a subject's compliance. It is the department's policy that officers should attempt to gain the voluntary compliance of subjects when possible. However, officers are

8. Nearly all officers who begin training graduate from the police academy with fewer than 3% of officers failing. The probationary period consists of eighteen months of active duty. Officers spend the first six months in the academy and the final twelve months in probationary district assignments. Time absent from duty does not apply toward completion of the probationary period.

9. A smaller share of officers work in specialized units, such as Canine, Marine/Helicopter, SWAT, or Bomb Squad units. Since these specialized units operate across geographic districts, we omit them from the analysis.

not required to take actions that endanger themselves or third parties.[10]

In general, the CPD requires officers to use force that is "objectively reasonable, necessary, and proportional" to the subject's actions (Chicago Police Department General Order G03-02, 2017). However, there is no formal definition of "objectively reasonable." The CPD instructs officers to consider whether there is an imminent threat to themselves or third parties, how much harm the threat poses, and whether the subject has immediate access to weapons. When assessing the validity of force, the CPD explicitly accounts for imperfect information regarding the suspect's compliance and that officers' make decisions quickly under tense circumstances.

The requirement of proportional force relies on the officer's contemporaneous beliefs about the threat she faces. These beliefs may differ from the suspect or the threat determined by an objective observer. The guidelines permit deadly force when the officer believes that the suspect poses the imminent threat of great bodily harm. Officers may also use deadly force when the suspect committed a forcible felony threatening the infliction of great bodily harm and attempted to avoid arrest. Under the guidelines, the department only permits officers to use this type of force as a last resort when all other de-escalation methods have failed.[11]

10. When attempting to gain the compliance of subjects, officers have several options available to them. Officers can use mitigation efforts such as verbal directions to gain compliance without using force. They may also use control tactics such as handcuffing or applying pressure to sensitive areas. Officers are also permitted to use higher-level responses with or without weapons; these include open hand strikes, punches, kicks, and other forms of physical violence. Lastly, the CPD permits officers to use Tasers, pepper spray, batons, and firearms under some circumstances.

11. Despite these guidelines, a 2015 probe by the Justice Department found that the CPD engaged in a "pattern or practice" of unconstitutional use of force. The Justice Department's report noted that CPD officers regularly engage in behavior that endangers themselves, resulting in unnecessary and avoidable force-use. It attributed the unconstitutional force to deficiencies in accountability and insufficient support for officer wellness and safety. The findings further noted that the unconstitutional force fell most on Black and Latino neighborhoods. See of Justice Press Release 15057 (2017) for more details.

1.3 Data and Summary Statistics

We use four sources of administrative data from the Chicago Police Department. Use of force and injury data come from the CPD's Tactical Response Reports (TRR) for non-juvenile suspects. The CPD requires that officers fill out Tactical Response Reports after an officer uses more than a minor level of force.[12] While minor levels of force do not require a TRR, officers must also fill out TRRs when a suspect alleges an injury, if the suspect resists arrest or in situations where the suspect uses physical violence (General Order G03-02-02).

Our data encompass over 16,000 instances of force by the CPD between January 1, 2005 and October 31, 2016. These data have numerous strengths relative to other existing data sets. They cover almost every instance of police use of force in Chicago, regardless of whether the officer injures or kills the suspect.[13] Second, the data contain detailed information about the time and location of the incident along with suspect, officer, and interaction characteristics. We supplement the data with officer employment records that include unit assignments and report the officer's start date, because it is critical to our identification strategy.[14]

To help identify the mechanisms driving the increase in force use, we supplement this data with data on complaints issued against officers and arrest data.[15] The complaint data contain all allegations of misconduct filed by civilians or other officers from January 1, 2005 until June 17, 2016, including the date of the incident and details of the actions resulting in a complaint. We match this data to the TRR data using the complainant's self-reported

12. This includes firearms, impact munitions, Tasers, acoustic devices, impact weapons, mechanical actions/techniques, or chemical weapons. Minor levels of force include things like holds or handcuffing. The police department also requires TRRs for force involving canines, but we exclude canine units from the analysis.

13. The data exclude incidents involving juveniles and subjects with unknown ages because juvenile records are not subject to Freedom of Information Act requests.

14. We exclude all police officers without a recorded start data from the analysis because we cannot link these officers to a police academy cohort.

15. See Ba (2017) for a detailed discussion of this data.

incident date. We investigate three specific types of complaints: force and verbal abuse, improper search or arrest, and failure to provide service. The arrest data contain all CPD arrests of adults during our sample period, including crime type, arrestee demographics, and arrest date and time. (see Section 1.8.1 for more details).

The data do have some limitations. While we observe the presence of any alleged injury to officers or civilians, we do not observe the nature, severity or cause of the injury.[16] We restrict our treatment definition to injuries that occur during interactions with suspects who allegedly attacked the officer.[17] Misclassification acts as measurement error and may lead us to consider control periods (no former peer was injured) as treated periods (at least one former peer injury), however this attenuates our effects as long as the covariance between former peer injuries and measurement error does not exceed the variance of the measurement error itself (Aigner, 1973; Black et al., 2000).

Lastly, CPD officer unit assignment data records officers as part of the "academy" unit until they finish their probationary period rather than graduate from the academy. We use the sample of officers who start the academy after January, 2001. We cannot observe the officers' geographic assignments in the year between graduation and the end of the probationary period. Since local non-compliance shocks constitute a significant threat to identification, we exclude every officers' probationary year from the analysis. We drop officers who leave the academy before six months or who graduate after our sample period.

A total of 4,429 officers start the academy between 2001 and 2013 (See Table 1.2). We drop 967 officers who do not enter into a geographic district after leaving the academy,

16. The CPD refused to provide this information in the FOIA request citing HIPPA privacy regulations.

17. The data does not include information about the nature or extent of officer injuries. However, Tiesman et al. (2019) reports that the most common cause of injuries-on-duty is violence. Most of these injuries were to the hands, legs, neck, head, or shoulders. About 40% of injuries were contusions, abrasions, lacerations, fractures, or dislocations. The other 60% were sprains, strains, or other. In their sample, assault-related injuries grew between 2003 to 2011. The Bureau of Labor Statistics also reports that of the 27,660 on-the-job injuries reported in their 2014 sample, violence caused 27% of injuries. The next most common category was falls, slips, and trips; this category accounted for 25.3% of injuries. Overexertion followed, accounting for 21.4% of injuries.

leaving us with 3,462 officers (See Table 1.3) and a total of 986,111 officer-week observations between 2004 and 2016 (the time period for which we observe TRRs). Of these officers, 2,836 use force at least once in the sample, with 2,000 instances accompanying an injury or alleged injury to the suspect. In our sample, 1,280 officers experience injuries. Nearly all officers (3,429) experience at least one injury to a member of their police academy cohort.

Table 1.2: Characteristics of All Officers who Start at the Police Academy

	count	mean	sd	min	max
Officer Male	4429	0.77	0.42	0	1
Officer Black	4429	0.17	0.37	0	1
Officer White	4429	0.49	0.50	0	1
Officer Hispanic	4429	0.34	0.47	0	1
Officer Age	4429	28.8	4.39	21	42

Note: This table reports descriptive statistics for all officers who enter the police academy during our sample period. Age at test is proxied by the officer's age at the most recent police exam.

Table 1.3: Characteristics of Officers who Enter Geographic Police Units

	count	mean	sd	min	max
Officer Male	3461	0.78	0.41	0	1
Officer Black	3461	0.17	0.37	0	1
Officer White	3461	0.49	0.50	0	1
Officer Hispanic	3461	0.34	0.47	0	1
Officer Age	3461	28.8	4.34	21	42

Note: Age is measured at the age of taking the entrance exam. Age at test is proxied by the officer's age at the most recent police exam.

Table 1.4 displays the summary statistics of events and outcomes by week. An individual officer is very unlikely to be injured in a given week. However, 90% of weeks involve at least one officer injury and 10% of weeks involve an injury to a former peer. Officers use force in about 1.8% of weeks. When they use force, they tend to use more than one type. Officers arrest suspects in about 26% of weeks and get complaints for their actions in about 1.6% of weeks.

Table 1.4: Frequency of Events and Outcomes

	Weeks	Mean	Standard Deviation	10th Percentile	50th Percentile	90th Percentile
Injuries:						
Officer Injured	675	0.003	0.002	0.000	0.002	0.005
Cohort Member Injured	675	0.112	0.078	0.005	0.100	0.227
Former Peer Injured	675	0.105	0.074	0.004	0.094	0.214
Any Office Injured	675	0.904	0.293	0.957	1.000	1.000
Force Use:						
Any Force Use	675	0.018	0.007	0.01	0.018	0.027
Control	675	0.011	0.005	0.005	0.010	0.017
Without Weapon	675	0.015	0.006	0.008	0.014	0.022
Nonlethal	675	0.001	0.002	0.000	0.001	0.003
Force mitigation	675	0.019	0.007	0.011	0.019	0.028
Baton	675	0.001	0.001	0.000	0.000	0.002
Taser	675	0.001	0.002	0.000	0.000	0.003
Firearm	675	0.000	0.001	0.000	0.000	0.001
Force other unknown	675	0.001	0.001	0.000	0.001	0.003
Injured Suspect	675	0.006	0.003	0.002	0.005	0.010
Arrests:						
Any Crime	520	0.257	0.175	0.000	0.341	0.429
Municipal Code	520	0.015	0.011	0.000	0.017	0.027
Traffic	520	0.032	0.022	0.000	0.04	0.055
Warrant	520	0.058	0.04	0.000	0.075	0.098
Drug Crime	520	0.082	0.062	0.000	0.101	0.155
Property Crime	520	0.042	0.03	0.000	0.051	0.075
Violent Crime	520	0.066	0.047	0.000	0.08	0.118
Other	520	0.083	0.058	0.000	0.102	0.147
Complaints:						
All Complaints	675	0.016	0.008	0.007	0.015	0.026
Force and Verbal	675	0.004	0.004	0.000	0.004	0.01
Arrest and Search	675	0.006	0.004	0.000	0.005	0.01
Failure to Provide Service	675	0.003	0.003	0.000	0.003	0.006
Unbecoming Conduct	675	0.000	0.001	0.000	0.000	0.001

Note: This table reports descriptive statistics for each of the weeks in the data set. Since officers are joining throughout the sample period, the composition of officers differs across weeks. The mean value represents the probability that each event or outcome occured at least once in the sample week.

1.3.1 Verifying Random Assignment to Cohort

We verify random assignment to cohort using the procedure from Guryan et al. (2009) to correct for the mechanical negative correlation between one's characteristic and that of its peer group.[18] In our case, this takes the form

$$X_{igr} = \pi_0 + \pi_1 \bar{X}_{-i,gr} + \phi \bar{X}_{-i,r} + \epsilon_{igr}. \tag{1.1}$$

Here, X_{igr} is the average pre-determined characteristic of officer i in academy cohort g chosen from test cohort r. We approximate the test cohort using cohorts which begin between two test dates.[19] The pre-determined characteristics we observe are the officer's sex, age at appointment, and race, and the descriptive statistics for these variables appears in Table 1.2.

We regress X_{igr} on $\bar{X}_{-i,gr}$, the class cohort leave-out-mean of X_{igr}, and $\bar{X}_{ir}$, the mean of X_{igr} for all individuals in test cohort r. Estimates of Equation 1.1 for each of the pre-determined observable is presented in Table 1.5. The lack of statistical significance on the class cohort leave-out-mean coefficients suggest that an officer's sex and race are randomly assigned. We find a positive correlation between an officer's age and the age of their class. However, this difference is economically small. A one-year increase in the average age of a class cohort is associated with a half a year increase in an officer's age. Overall, these tests suggest that there is random assignment to police academy cohorts in practice. The discrepancy in the age variable may be due to a preference granted to Veterans in the lottery process.

18. Guryan et al. (2009) corrects the procedure in Sacerdote (2001) to show that there is a mechanical negative correlation between one's characteristics and that of its peer group arising from the fact that peers are, in a sense, sampled without replacement. That is, peers with high values of a characteristic are chosen from a group with slightly lower mean characteristics than those with low ability because removing them from the group reduces the mean of the set. They also show that this bias is decreasing in the size of the set. Randomization in our setting occurs on the entrance-exam cohort level. Therefore, the randomization sets we consider range from 52 to 1,300 officers (see Table 1.1)

19. We do not know for certain whether officers in two cohorts come from the same test cohort as the CPD does not track this information (see Section 1.8.7 for more details). In our research design, we overcome this issue by using officer level fixed effects.

Table 1.5: Testing the Random Assignment of Police Entrance Lotteries

	(1) Officer Male	(2) Officer Black	(3) Officer White	(4) Officer Hispanic	(5) Officer Age
$\bar{X}_{-i,gr}$	0.224	0.116	0.0387	-0.0502	0.523**
	(0.124)	(0.112)	(0.125)	(0.141)	(0.165)
Constant	86.79*	25.60**	60.34*	37.76*	2610.3*
	(42.50)	(9.946)	(28.23)	(18.78)	(1409.4)
$\bar{X}_{-i,r}$	YES	YES	YES	YES	YES
R-squared	0.346	0.423	0.351	0.329	0.342
Observations	3468	3468	3468	3468	3468

Note: Table includes results from estimating Equation 1.1 with various predetermined officer-level characteristics. Sample includes every officer who started at the police academy between January 2002 and December 2013 regardless of whether they work in a geographic unit after graduation. Officer age is measured at the age of taking the entrance exam. Standard errors are in parenthesis and are clustered by the test cohort. We assign officers to a test academy cohort based on the last test that occur ed before they began at the police academy. However, we do not observe the officer's actual test date.

1.4 Research Design

The empirical analysis aims to identify the causal effect of an injury-on-duty in one's network. There are five primary challenges to identifying this effect: (i) observing the officer's reference group, (ii) correlated effects, wherein officers who are more likely to get into altercations with suspects are also more likely to be in the same network, (iii) common shocks to civilian non-compliance, (iv) simultaneity between a civilian's non-compliance and an officer's decision to use force, and (v) simultaneity between an individual's actions and their peer's action (known as the reflection problem).

The ideal research design for overcoming these challenges uses two-stages of randomization: individuals are first randomized into a group, then, within a group, individuals are further randomly selected into treatment or control conditions (Philipson, 2000; Hirano and Hahn, 2010; Angrist, 2014; Baird et al., 2018). We approximate this design by choosing to define our peer group as officers who, through the random lottery, attended the police academy together and now work in a different district, then we exploit the quasi-random timing of force-related injuries to those "former" peers.

In principle, the relevant network is the set of officers whose injury status is observable to the officer in question, which could all officers or officers in the same unit for example. Following Ager et al. (2018), we consider one's network to be their former peers, i.e., officers who attended the police academy together but now work in a different district. Henceforth, we adopt their language and refer to these peers as former peers to differentiate them from members of the police academy who still work in the same district.[20]

Our definition helps alleviate concerns that we observe correlation in officer injuries and force-use because officers who are more likely to get into altercations with suspects are also more likely to be in the same group. More aggressive officers sorting into more aggressive networks could spuriously generate a correlation between injuries and use of force that could be mistaken for a peer effects in the data. The CPD recruitment process ensures that each group's unobservable characteristics, such as the officers' aggression, are balanced across academy cohorts within each test cohort.

We allow for interference between officers within a district, but not across districts. Our definition also addresses common shocks to civilian non-compliance. When some shock reduces the probability that civilians comply with the officer's requests, then both the risk to the officer and the returns to using force will increase. Using former peers allows us to rule out district-level shocks to civilian non-compliance because we compare treated officers to other officers who face the same non-compliance rate.

Finally, we use the panel structure of the data in conjunction with a difference-in-differences design to eliminate bias arising from both types of simultaneity. We use lagged peer injuries because contemporaneous peer injuries may be confounded with some state of the world that affected the non-compliance rate for all parties in that week, and it may also take time for officers to learn about the injuries of their peers. Moreover, Angrist (2014) shows that designs relying on random variation in cohort assignments do not overcome the

20. This means that we consider individuals to be untreated if a member of their academy cohort who currently works in the same district is injured.

reflection problem because identification relies on finite-sample fluctuations in treatment assignment. As such, we use the quasi-random timing of injuries to former peers as an approximation of the random assignment of subjects to treatment that is independent of the assignment of groups.

In practice, we construct the counterfactual outcomes within district d and week t, using injuries to former peers as the treatment. We identify the effect of an officer injury using an event study that compares individuals experiencing and not experiencing an injury to a former peer and combining them into a difference-in-differences estimator. The event of a former peer's injury occurs at time $t = E_i$ for individual i. We denote individual fixed effects as λ_i and district-week fixed effects as λ_{dt}. The primary equation used to recover the causal effect of peer injuries is

$$Y_{idgt} = \lambda_i + \lambda_{dt} + \beta \cdot \mathbb{1}[t = E_{g,-d} + 1] + \epsilon_{idgt}. \tag{1.2}$$

Here, the unit of observation is an officer-week. The outcome variable, Y_{idgt}, is an indicator function, equal to one if the outcome is realized for officer i working in district d during week t who belongs to academy cohort g. For example, in our main specifications, Y_{idgt} is equal to one if officer i chose to use force in week t and zero otherwise. An advantage of this approach is that using the prevalence of the outcome allows us to avoid endogeneity issues from conditioning on interactions with police.[21] The treatment, $\mathbb{1}[t = E_{g,-d} + 1]$, is an indicator equal to 1 if a former peer (an officer who attended the police academy with officer i and is working in a different district) was injured in the previous week.

The individual fixed effects, λ_i, account for time-invariant individual-level differences in the outcome. For example, they account for time-invariant differences in how each officer interprets a suspect's actions as non-compliance. They also subsume test-cohort fixed effects which is the level of randomization to group.

21. See Fryer (2018), Heckman and Durlauf (2020), and Fryer (2020) for discussions of this issue.

District-week fixed effects, λ_{dt}, account for district-week level differences in the costs or benefits of choosing $Y_{idgt} = 1$. These fixed effects control for district-specific shocks to civilian or officer aggression, such as the weather or pollution (Annan-Phan and Ba, 2020; Herrnstadt et al., 2016) and control for common shocks under the partial interference assumption.

The coefficient of interest, β, estimates the change in the outcome for affected officers relative to officers in the same district who did not experience a former peer injury in the previous week. Standard errors are clustered on the academy cohort level to allow for arbitrary correlation of errors within each of the 81 cohorts. The main identifying assumption is that the change in the outcome in a given district-week is independent of whether an injured officer started the police academy in the same month as the officer.

To assess this assumption's plausibility and examine the dynamic effects of a peer injury, we regress the outcomes on lags and leads of injuries to a former peer. We denote event time in this regression as τ. We omit the dummy for the week before a former peer is injured so that we can interpret the coefficients relative to the week before injury. We set period -6 to be equal to one when the event was six or more weeks before the injury and set period 6 to be one if the week is six or more periods after the injury.[22]

$$Y_{idgt} = \lambda_i + \lambda_{dt} + \sum_{\tau} \beta_\tau \cdot \mathbb{1}[t = E_{g,-d} + \tau] + \epsilon_{idgt}. \tag{1.3}$$

where $\tau = \{-6+, -5, -4, -3, -2, 0, 1, 3, 4, 5, 6+\}$. In this regression, the coefficients of interest are now β_τ. These coefficients estimate the change in the outcome between period $t = -1$ and τ for officers who experienced a peer injury relative to members of the same district who did not. Insignificant β_τ estimates before the event alleviate concerns that the groups differ in the probability of encountering non-compliant civilians or signal interpreta-

22. The lags and leads will also alter the composition of individual-weeks that we observe. The first and last five weeks of every individual's observations in the panel will be excluded from the regression since they are not observed with either five lags or five leads.

tion; however, unobserved post-treatment shocks specific to members of a particular police academy cohort may still threaten the identification of β_τ.

Officers experience multiple events over their time in our sample. Standard event studies usually include one event per cross-sectional unit and mutually exclusive dummy variables representing periods before and after treatment. Our setting departs from this standard. While the probability an individual officer gets injured is one quarter of one percent, over 95% of weeks in our sample contain at least one officer injury, and officers have roughly a 1 in 8 chance of experiencing a peer injury each week. Over the observed portion of an officer's career, the average officer experiences 0.89 injuries, 43.62 injuries to former peers, and 368.48 injuries to any police officer. Thus, β_τ can represent the effect for a period, which is both a pre-treatment period and a post-treatment period. Assuming the response to treatment does not vary based on the number of previous events, this will bias the pre-trend estimates away from zero and make it more likely for us to find significant pre-trends. However, in nearly all specifications, we do not find any evidence of significant pre-trends.

There is no accepted method of conducting event studies when there are multiple or overlapping events. The Monte Carlo simulation results in Sandler and Sandler (2014) suggest that allowing multiple event dummies to be non-zero at one time produces unbiased results under a similar data generating process. Further, they show that restricting the estimation to consider only a single event or using only periods that have a single event per individual/event/time produces biased estimates of the treatment effect. We follow their guidance in our estimation and allow multiple event dummies to be non-zero.

1.5 Results

This section demonstrates that after an former peer is injured during a violent arrest, officers substatially increase their propensity to use force in a short time period. Officers are also more likely to injure suspects and receive complaints about their conduct during the first two weeks after an injury to a former peer. Officer tenure reduces the magnitude of these effects. Moreover, the effects are moderated by social distance, as officers respond twice as strongly to the injury of former peers of the same race.

1.5.1 *Propensity to Use Force and Injure Suspects*

We first consider whether officers increase their propensity to use force after a former peer's injury, to avoid the reflection problem (Manski, 1993) and local shocks to civilian non-compliance. Table 1.6 displays the results of estimating Equation 1.2 with the outcome being the use of force and each column containing different controls and specifications. Column (1) displays a strong correlation between officer injuries and the propensity to use force. Former peer injuries are associated with a 24% higher likelihood of using force in the following week.

Table 1.6: Effect of Injuries to Former Peers on the Propensity to use Force

	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force
Former peer in previous week	0.00380***	0.00150**	0.00134**	0.00127**	0.0747***
	(0.000782)	(0.000582)	(0.000609)	(0.000556)	(0.0270)
Constant	0.0178***				
	(0.000585)				
Model	OLS	OLS	OLS	OLS	Poisson
Percent Increase	21.28	8.43	7.48	7.09	7.47
Test Fixed Effects	NO	YES	YES	YES	YES
Unit-Week Fixed Effects	NO	YES	YES	NO	NO
Number of Former Peers	NO	NO	NO	YES	YES
Individual Fixed Effects	NO	NO	YES	NO	NO
Pre-trend Test	0.000	0.799	0.819	0.822	0.594
R-squared	0.000	0.025	0.025	0.042	
Observations	896363	896250	896250	896250	618513

Note: Column 1 displays estimates from a linear regression of an indicator for any force used by the officer on the first lag of injuries to former peers. Column 2 controls for unit-week fixed effects. Column 3 controls for unit-week and number of former peer fixed effects. Column 4 controls for unit-week, number of former peers, and estimated test period fixed effects. Column 5 estimates Equation 1.2, controlling for individual and unit-week fixed effects. Column 6 Equation 1.2 using Poisson maximum likelihood estimation. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Figure 1.1 displays all percent changes from Column (5) in this regression testing for differences in pre-trends. Column (1) in Appendix Table 1.17 displays all of the coefficients on lags and leads for Column (6) in this table. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

However, to establish a causal relationship, we must address three main threats to identification. First, we account for local time-varying shocks to civilian non-compliance by adding district-week fixed effects in Column (2). Second, we account for heterogeneity in class-size that may have affected the quality of teaching and the probability of experiencing an event in Column (3). Next, we account for sorting into more aggressive networks using our predicted test-period fixed effects in Column (4).[23] Finally, in Column (5) of Table 1.6, we replace test fixed effects with individual fixed effects. These fixed effects allow us to control for time-invariant differences in an individual's propensity to use force and overcomes any mis-classification of test dates.[24] We find that the treatment effects do not change substantially or qualitatively with the introduction of individual fixed effects.

Estimates in Columns (4) and (5) are nearly identical, with a former peer's injury increasing an officer's likelihood of using force in the following week by 7% relative to the baseline rate. Finally, because the outcome is rare (The probability an officer uses force is under 2% per week), we include a Poisson specification as a robustness check in Column (6).[25] The coefficient in Column (6) of Table 1.6 can be readily interpreted as an approximation of the percent change. We find this coefficient is similar to the OLS estimate in size and significance.

To better understand the consequences of the increases in force-use, we estimate Equation 1.2 with use of force resulting in a suspect as the outcome variable in Table 1.7, with the same controls and specifications as in Table 1.6. Column (1) of Table 1.7 shows that the baseline rate of suspect injury per week is 0.55%. Based on Column (5) Injuries to former peers increase an officer's propensity to injure suspects by 10% of this baseline mean. As in

23. Changes in the applicant pool may lead to cohorts that systematically differ from one another on unobservable differences that make them both more likely to experience injuries and inflict force.

24. One's test cohort is time-invariant and is therefore perfectly collinear with the individual fixed effect.

25. We choose to estimate the model using Poisson maximum likelihood estimation rather than with a Logit model because of the incidental parameters problem (Neyman and Scott, 1948; Hahn and Newey, 2004). Wooldridge (1999) shows that Poisson maximum likelihood works well with a binary response variable.

Table 1.6, comparing Columns (4)-(6) shows that these results are similar when replacing individual-level fixed effects with test-period fixed effects or using a Poisson regression.

Table 1.7: Effect of Past Peer Injuries on Suspect Injuries

	(1) Injure Suspect	(2) Injure Suspect	(3) Injure Suspect	(4) Injure Suspect	(5) Injure Suspect	(6) Injure Suspect
Former peer in previous week	0.00140***	0.00122***	0.000744**	0.000540*	0.000531*	0.0930**
	(0.000349)	(0.000358)	(0.000318)	(0.000310)	(0.000301)	(0.0438)
Constant	0.00528***	0.00530***	0.00535***	0.00537***	0.00537***	-3.211***
	(0.000218)	(0.000212)	(0.000179)	(0.000149)	(0.0000340)	(0.00607)
Model	OLS	OLS	OLS	OLS	OLS	Poisson
Percent Increase	26.49	23.05	14.1	10.24	10.06	9.3
Test Period Fixed Effects	NO	YES	YES	YES	YES	YES
Unit-Week Fixed Effects	NO	NO	NO	YES	NO	NO
Number of Former Peers	NO	NO	NO	NO	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	NO	NO
Pre-trend Test	0.000	0.002	0.032	0.086	0.085	0.097
R-squared	0.000	0.021	0.022	0.022	0.031	
Observations	986,111	986,088	986,088	986,088	986,088	233,326

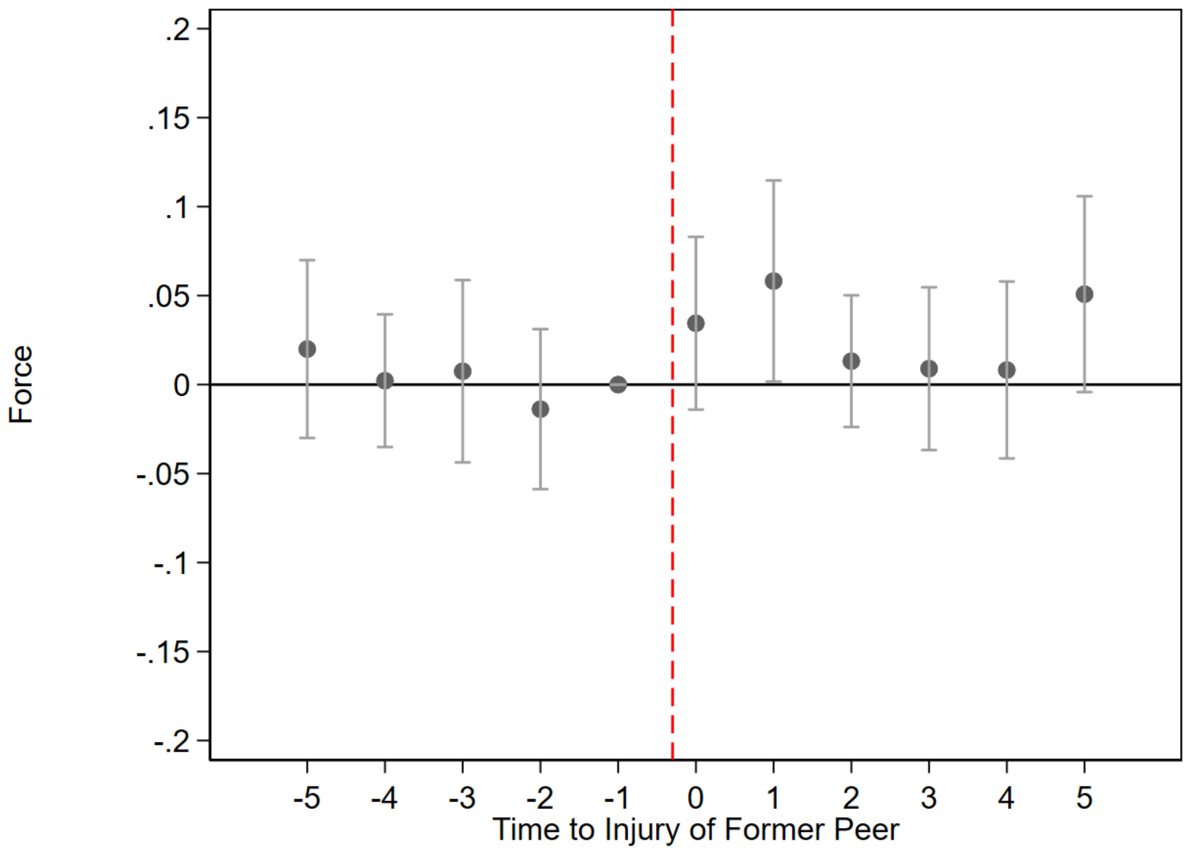
Note: Column 1 displays estimates from a linear regression of an indicator for a suspect injury on the first lag of injuries to former peers. Column 2 controls for unit-week fixed effects. Column 3 controls for unit-week and number of former peer fixed effects. Column 4 controls for unit-week, number of former peers, and estimated test period fixed effects. Column 5 estimates Equation 1.2, controlling for individual and unit-week fixed effects. Column 6 Equation 1.2 using Poisson maximum likelihood estimation. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Figure 1.7 displays all percent changes from Column (5) in this regression testing for differences in pre-trends. Column (2) in Appendix Table 1.17 displays all of the coefficients on lags and leads for Column (6) in this table. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

1.5.2 Event Study

Next, we investigate the dynamics of the treatment effect and test whether there are parallel pre-trends between officers with and without a former peer injured. To do so, we estimate Equation 1.3 where the outcome variable is any force used in the week after a former peer is injured.[26] We estimate the base rate of force as the constant term from a regression of Equation 1.3 without individual or unit-week fixed effects. Figure 1.1 displays the coefficient estimates from Equation 1.3 divided by the base rate. This allows us to interpret the effects

26. As cohorts join throughout the sample period, the estimation equation will drop individuals who are not employed for five weeks before and after an exposure. As a robustness check, Column (1) in Appendix Table 1.17 repeats this exercise using a Poisson regression specification.

Figure 1.1: The Effect of Past Peer Injuries on Police Use of Force



Note: The graph shows coefficient estimates using Equation 1.3 divided by baseline rate of force use and 95% confidence intervals. The baseline rate of force is calculated as the constant term from a regression of force on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer. Red vertical line represents the injury week.

as percent changes from the baseline propensity to use force.

The effects of a former peer injury in the weeks before the injury are neither economically nor statistically significant. Baseline use of force is also small, with 1.78% of officers using any force in a given week. In the week of a former peer injury, the use of force increases by around 3% of the baseline mean. We view this period as partially treated as some officers experience injuries toward the beginning of the week, while the weeks after the injury are fully treated as injuries may happen toward the end of the week and it may take a few days for officers to learn of the injury.

In the week after a former peer is injured, officers increase their use of force by roughly 7% of the baseline mean. The effects quickly dissipate, immediately losing significance the first-week after exposure. This pattern is consistent with Card and Dahl (2011) and Munyo and Rossi (2015), who both find that emotional shocks have a short-lived effect on violence.

1.5.3 Types of Force Used by Officers

Next, we investigate what types of force officers used to respond to the injury of former peers. The CPD use-of-force model governs the choice of which type to use, and, in general, the more resistance the officer faces, the more force they are permitted to use.

The lowest level of force is called control tactics. It includes actions such as escort holds, wrist locks, emergency handcuffing, and armbars. Above that is physical strikes, such as takedowns, open hand strikes, punches, kicks, or elbows that do not involve more than the officer's body. The CPD classifies force involving weapons as non-lethal if it involves a chemical weapon and classifies other force involving a weapon by weapon type (i.e. baton or impact weapon, Taser, or firearm). We categorize all other uncommon types of force as "Other".

We estimate Equation 1.2 on indicators for using each type of force separately and present the results in Table 1.8. The majority of instances of force recorded in this data are from force with no weapon, control tactics, and Tasers, and the rarest type of force is the use of firearms, followed by impact weapons such as batons. Note that there is likely much lower under-reporting for types of force that are harder to conceal (firearms and Tasers) because of the CPD regulations against it.

Similar to our main results, we do not find evidence of pre-trends in any specification except for non-lethal force. We find that the officers primarily respond to a former peer's injury by increasing control tactics and force without weapons, by about 8% percent relative to baseline for each. There is a substantial increase in officers using a firearm in the week

Table 1.8: Heterogeneous Effects by Type of Force

	(1) Control	(2) No Weapon	(3) Non-Lethal	(4) Baton	(5) Taser	(6) Firearm	(7) Other
Former peer in previous week	0.000715**	0.000951*	-0.0000165	-0.0000373	0.0000913	0.000116*	0.0000689
	(0.000348)	(0.000480)	(0.0000795)	(0.0000653)	(0.000131)	(0.0000639)	(0.000153)
Constant	0.0103***	0.0144***	0.000865***	0.000498***	0.00167***	0.000274***	0.00108***
	(0.0000392)	(0.0000541)	(0.00000896)	(0.00000736)	(0.0000147)	(0.00000720)	(0.0000173)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Percent Increase	8.19	7.74	-1.92	-9.17	7	48.36	7.17
Unit-Week Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Pre-trend Test	0.684	0.742	0.010	0.407	0.495	0.923	0.545
R-squared	0.034	0.039	0.029	0.022	0.024	0.022	0.023
Observations	986,088	986,088	986,088	986,088	986,088	986,088	986,088

Note: Columns 1 through 7 display coefficients from estimates of Equation 1.2 where the outcome variable is an indicator representing whether the officer used a specific type of force. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Appendix Table 1.18 displays all coefficients from this regression testing for differences in pre-trends. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

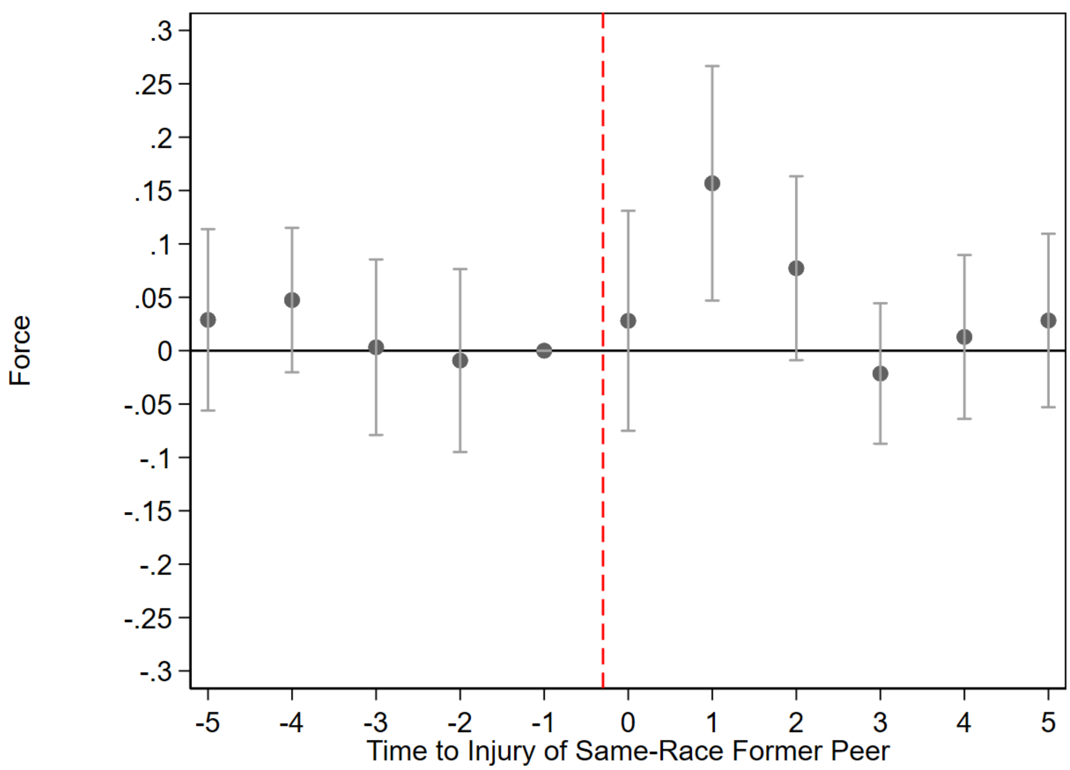
after a former peer is injured, 48% relative to baseline; however, this represents a very small percentage point increase due to the rarity of firearm usage. Assuming that officers are using force in alignment with the CPD use-of-force model, the increase in force use is primarily driven by encounters with low-resistance suspects, suggesting that officers would not have deemed these suspects to be a risk had their peer not been injured in the previous week.

1.5.4 Same-Race Former Peers

Our definition of former peer, thus far, has included everyone who attended the academy together and now works in a different police district. For peer injuries to have any effect, these individuals must have been acquainted and maintained their bonds after the academy ended. Given that even randomly assigned groups still produce homophilic friendships (Carrell et al., 2013), we expect stronger bonds to be between individuals of the same race (Marmaros and Sacerdote, 2006). Such homophily will attenuate effects, as we will be pooling strongly and weakly treated officers. Following McPherson et al. (2001), we assume individuals of the same race who attended the academy together are more likely to be a part of the same network. We perform the same analysis as before, but we now define a "former" peer group as members of the same academy cohort now working in different districts who are all same race.

Figure 1.2 repeats our analysis from the previous section using this definition of peer groups. We find that officers respond twice as strongly to the injury of a same-race former peer. Officers increase their propensity to use force by 16.51% in the week after a former peer is injured with a similar baseline probability of using force. This result is in line with past literature on peer effects that shows peer effects mainly operate within race (Garlick, 2018).

Figure 1.2: The Effect of Same Race Past Peer Injuries on Force Use



Note: The graph shows coefficient estimates using Equation 1.3 divided by baseline rate of force use and 95% confidence intervals. The baseline rate of force is calculated as the constant term from a regression of force on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer of the same race. Red vertical line represents the injury week.

1.5.5 *Injuries Unlikely to be Caused by Suspects*

As a robustness check, we investigate officers' propensity to respond to injuries to former peers that occurred during interactions with suspects who displayed low levels of resistance. That is, when the former peer reported that the suspect did not attack them during the encounter. While we do not know the cause of these injuries, Tiesman et al. (2019)'s results suggest that these may be due to falls, slips, or trips. If officers respond to a perceived threat, we would expect officers to not respond to these types of injuries.

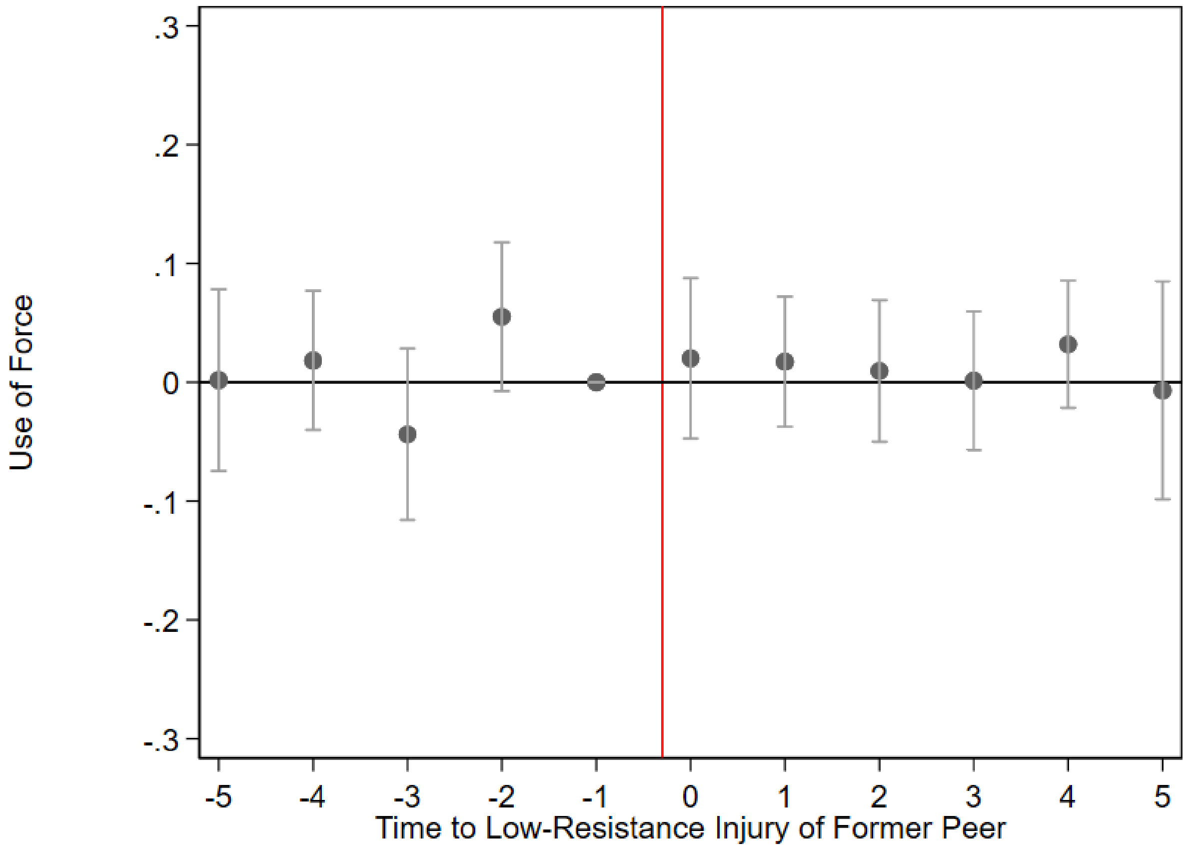
Table 1.9 and Figure 1.3 display estimates of the effect of these injuries on an officer's propensity to use force. In line with our expectations, we find no evidence that officers respond to injuries of former peers that unlikely to be caused by suspects.

Table 1.9: Effect of Low-Resistance Injuries to Former Peers on Force Use

	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force	(6) Force
Former peer in previous week	0.00335***	0.00238***	0.000895	0.000479	0.000366	0.0161
	(0.000763)	(0.000901)	(0.000678)	(0.000609)	(0.000565)	(0.0286)
Constant	0.0175***	0.0175***	0.0176***	0.0176***	0.0177***	-3.006***
	(0.000591)	(0.000575)	(0.000440)	(0.000355)	(0.0000341)	(0.00204)
Model	OLS	OLS	OLS	OLS	OLS	Poisson
Percent Increase	19.18	13.64	5.12	2.74	2.1	1.61
Unit-Week Fixed Effects	NO	YES	YES	YES	YES	YES
Number of Former Peers	NO	NO	YES	YES	NO	NO
Test Period Fixed Effects	NO	NO	NO	YES	NO	NO
Individual Fixed Effects	NO	NO	NO	NO	YES	YES
Pre-trend Test	0	0.002	0.034	0.068	0.079	0.101
R-squared	0.000	0.022	0.023	0.024	0.040	
Observations	986,111	986,088	986,088	986,088	986,088	607,688

Note: Column 1 displays estimates from a linear regression of an indicator for any force used by the officer on the first lag of injuries to former peers. In this table, the definition of injury only uses injuries to officers for which the officer reported that the suspect did not use physical resistance. Column 2 controls for unit-week fixed effects. Column 3 controls for unit-week and number of former peer fixed effects. Column 4 controls for unit-week, number of former peers, and estimated test period fixed effects. Column 5 estimates Equation 1.2, controlling for individual and unit-week fixed effects. Column 6 Equation 1.2 using Poisson maximum likelihood estimation. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Figure 1.1 displays all percent changes from Column (5) in this regression testing for differences in pre-trends. Column (1) in Appendix Table 1.19 displays all of the coefficients on lags and leads for Column (6) in this table. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Figure 1.3: Effect of Low-Resistance Injuries to Former Peers on Force Use



Note: The graph shows coefficient estimates using Equation 1.3 divided by baseline rate of force use and 95% confidence intervals. Events are injuries to former peers that occurred during interactions where the officer reported that the suspect used no or non-violent resistance. The baseline rate of force is calculated as the constant term from a regression of force on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer. Red vertical line represents the injury week.

1.5.6 *Complaints*

Next, we investigate the effect of injuries to former peers on allegations of misconduct by the officers. When a police officer acts outside the confines of the US constitution or other relevant regulations, there are limited ways to detect this violation. Any civilian who believes an officer violated their Constitutional rights may sue the law enforcement agency or report the violation to an oversight agency. Complaints are generally easier to file than lawsuits and are the most readily available form of civilian feedback a police department can access (Ba, 2018; Ba and Rivera, 2019).

In this section, we consider whether officers are more likely to act in a way that generates a complaint in the week following former peer's injury. Table 1.10 and Figure 1.4 show estimates of Equations 1.3 and 1.2 respectively with an indicator for having any complaint as the outcome. We find that in the week following a former peer injury, officers are 7% more likely to engage in behavior that leads to a complaint of any type (Column (1)).

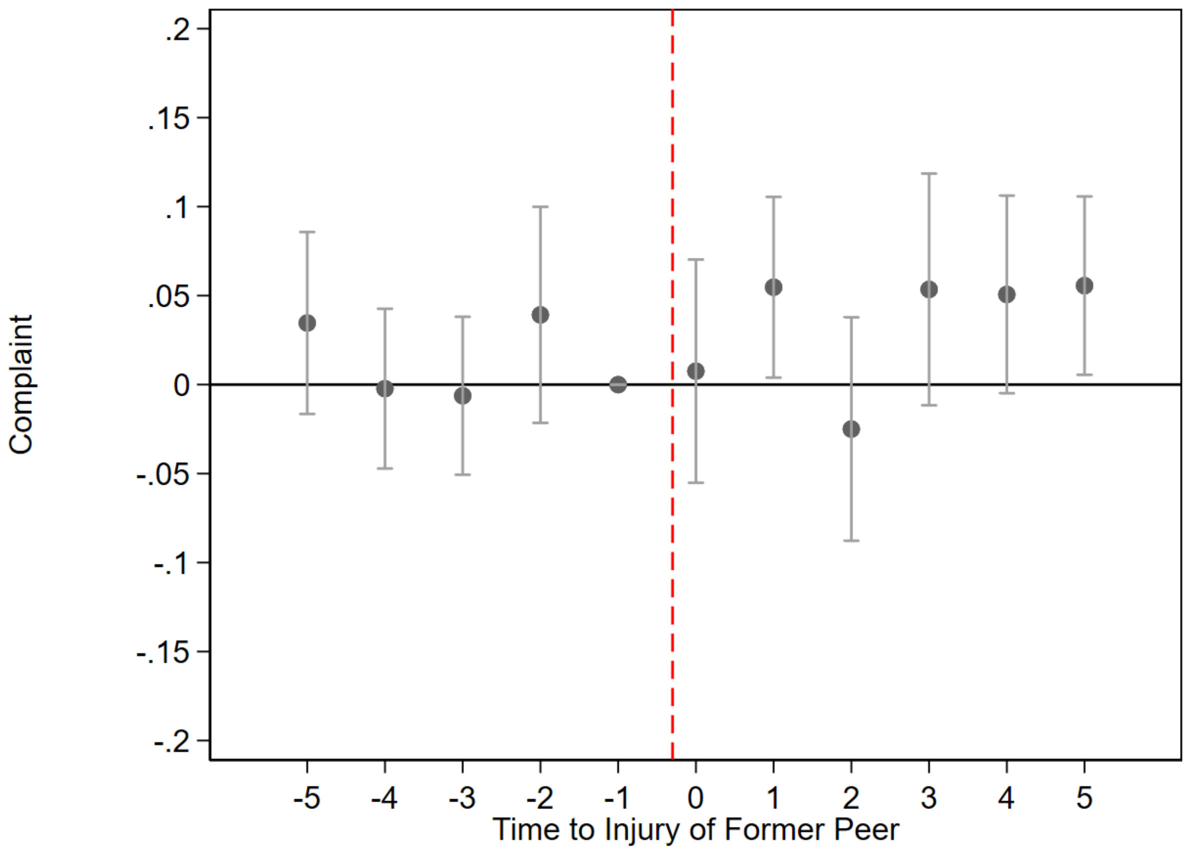
Table 1.10: Effect of Past Peer Injuries on Complaints Against Officers

	(1) All Complaints	(2) Force and Verbal	(3) Arrest and Search	(4) Failure to Provide Service	(5) Unbecoming Conduct
Former peer in previous week	0.000833**	-0.000134	0.000483**	0.000359**	0.00000963
	(0.000388)	(0.000201)	(0.000198)	(0.000179)	(0.0000577)
Constant	0.0131***	0.00339***	0.00496***	0.00250***	0.000255***
	(0.0000437)	(0.0000227)	(0.0000224)	(0.0000202)	(0.00000651)
Percent Increase	6.95	-4.39	10.73	15.69	3.52
Unit-Week Fixed Effects	YES	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES	YES
Pre-trend Test	0.412	0.516	0.117	0.468	0.550
R-squared	0.039	0.033	0.037	0.027	0.027
Observations	986,088	986,088	986,088	986,088	986,088

Note: Columns 1 through 5 display coefficients from estimates of Equation 1.2 where the outcome variable is an indicator representing types of complaints against the officer. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. Appendix Table 1.20 displays all coefficients from this regression testing for differences in pre-trends.

We then consider four types of complaints, and present estimates of Equation 1.2 with the outcome being an indicator for each type of complaint in Columns (2) through (5) of Table 1.10. The first two, excessive force or verbal abuse (Column (2)) and improper search or

Figure 1.4: The Effect of Past Peer Injuries on Complaints



Note:The graph shows coefficient estimates using Equation 1.3 divided by baseline rate of complaints and 95% confidence intervals. The baseline rate of complaints is calculated as the constant term from a regression of complaints on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer. Red vertical line represents the injury week.

arrest (Column (3)), constitute violations of an individual's constitutional rights during an officer's attempt to enforce the law. We also consider the failure to provide a service (FPS) complaints (Column (4)) and unbecoming conduct complaints (Column (5)). Analyzing different types of complaints enables us to distinguish between neglecting a civilian who desired help (for example, a potential victim of a crime), which would be an FPS complaint, and violations of a civilian's constitutional rights. Finally, complaints about unbecoming conduct cover actions like providing a false statement, being drunk and disorderly in public or on base, or insulting another officer.

We find no change in the use-of-force or verbal complaints following a former peer injury. However, complaints of a false arrest or improper search increase by nearly 11% in the week following a peer injury. This result is consistent with officers increasing the rate at which they violate a civilian's constitutional rights after a former peer is injured. We also find evidence that police are much more likely to neglect civilians requesting help. Complaints for the failure to provide service increase by 15.69% in the week following a peer injury. We do not find any evidence of additional complaints about unbecoming conduct.

1.5.7 Heterogeneity

To better understand how officers respond to peer injuries, we investigate heterogeneity based on officer tenure, the number of past events and suspect characteristics.

1.5.8 Effects of Officer Experience

We consider whether professional experience lessens the social influences in force use. Guryan et al. (2009) suggests that professional experience attenuates social influences such as peer effects. As officers gain more experience on the job, they learn how to avoid responding to their former peers' injuries. Similarly, there is a large experimental literature suggesting that market experience contributes to individual rationality (List, 2003; List et al., 2005;

List, 2011; Tong et al., 2016). This suggests that officers might be less likely to respond to emotional shocks as they gain experience.

Table 1.11 displays the effect of former peer injuries as well as an interaction term between former peer injuries and the officer's tenure in months since they started at the police academy for the outcomes consider so far. More experienced officers are less responsive to former peers' injuries in terms of using force, injuring suspects, or acting in a manner that causes a civilian to issue a complaint. This finding is consistent with the evidence on experience and rationality and the general finding that social influence decreases with experience. With respect to policing, this finding is also consistent with those in (Ta et al., 2021) who use body camera data to show that a police officer's emotional reactivity is lower in more experienced police officers.

Table 1.11: Heterogeneous Effects by Tenure

	(1) Force	(2) Injure Suspect	(3) Arrest	(4) Officer Injured	(5) Complaint
Former peer in previous week × Tenure (months)	-0.0000437**	-0.0000213**	0.000124*	0.00000149	-0.0000250**
	(0.0000185)	(0.0000106)	(0.0000625)	(0.00000546)	(0.0000115)
Former peer in previous week	0.00375***	0.00175**	-0.00475	-0.000254	0.00227***
	(0.00136)	(0.000779)	(0.00447)	(0.000378)	(0.000777)
Constant	0.0175***	0.00537***	0.299***	0.00236***	0.0131***
	(0.0000613)	(0.0000343)	(0.000242)	(0.0000188)	(0.0000436)
Model	OLS	OLS	OLS	OLS	OLS
Unit-Week Fixed Effects	YES	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES	YES
Pre-trend Test	NO	NO	NO	NO	NO
R-squared	0.742	0.159	0.945	0.07	0.491
Observations	0.040	0.031	0.280	0.026	0.039
N	986,088	986,088	933,782	986,088	986,088

Note: Columns 1 through 5 display coefficients from estimates of Equation 1.2 with various indicators and an interaction term between a lagged injury to a former peer and the officer tenure. Officer tenure is a continuous variable representing the number of months since the officer started at the police academy. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

1.5.9 Suspect Characteristics

Next, we investigate heterogeneity based on suspects' race by estimating equation 1.3 with the outcome being the use of force against a suspect of a particular race. We display the results in Table 1.12. We find that officers are significantly more likely to use force against African-American suspects. There are no significant increases in the probability of using force against white or Hispanic suspects. Although, the use of force against Hispanic suspects is imprecisely estimated.

Readers should use caution when interpreting these results. Roughly 81% of force uses and 80% of officer injuries result from interactions with African American suspects. As such, our results may be driven by the relatively small number of events observed for White and Hispanic suspects.

Table 1.12: Heterogeneous Effects by Suspect Characteristics

	(1) White Suspect	(2) Minority Suspect	(3) Male Suspect	(4) Female Suspect
Former peer in previous week	0.0000776	0.00114**	0.00115**	0.0000866
	(0.000118)	(0.000494)	(0.000558)	(0.000131)
Constant	0.00112***	0.0162***	0.0151***	0.00251***
	(0.0000133)	(0.0000557)	(0.0000630)	(0.0000147)
Model	OLS	OLS	OLS	OLS
Percent Increase	6.88	7.2	7.78	3.53
Unit-Week Fixed Effects	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES
Pre-trend Test	0.697	0.834	0.817	0.985
R-squared	0.036	0.039	0.039	0.026
Observations	986,088	986,088	986,088	986,088

Note: Columns 1 through 3 display coefficients from estimates of Equation 1.2 where the outcome variable is an indicator representing whether the officer used a specific type of force against a suspect of a given race. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Appendix Table 1.21 displays all coefficients from this regression testing for differences in pre-trends. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

1.6 Mechanisms and Alternative Interpretations

Having established that police officers respond to peer injuries, we now attempt to understand what mechanisms might be driving this behavior. We begin by ruling out potentially confounding effects that would challenge our interpretation of the main finding. These include officers mimicking the force-use of their peers through traditional peer effects and officers increasing their effort after a peer is injured. Then, we investigate whether officers are responding to peer injuries because these injuries provide some information about their injury risk or because of transitory emotional responses.

1.6.1 Officers Mimicking Peer Force Use

First, we investigate whether officers are actually responding to their former peer's decision to use force rather than their former peer's injury itself. A large body of work shows that the actions of one's peers influence decision making (Brock and Durlauf, 2001). For example, Murphy (2019) finds that misconduct by soldiers in the US Army tends to occur at similar times as the misconduct of peers. In our setting, a similar effect would be officers increasing the propensity to use force because they witness a peer doing so. This effect could potentially confound our result because, in 94% of instances where officers were injured, they also used force against the suspect (see Table 1.13).

Table 1.13: Force Use and Injuries

	Did not use Force	Used Force	Total
Not Injured	972,306	15,291	987,597
Injured	125	2,186	2,311
Total	972,431	17,477	989,908

Note:This table displays the frequency of force-use and injuries for every officer-week observation in our sample.

We investigate this potential confound using over 14,000 instances of force unaccompanied by an officer injury. We use these instances to investigate whether force-use mimicry is driving these results by estimating equations 1.2 and 1.3 with the outcome being force use and the treatment being an officer's former peer using force.

Column (1) of Table 1.14 shows that there is a strong correlation between an officer's use of force and the force use of former peers that was unaccompanied by an officer injury. However, after we control for individual and district-week fixed effects, we find no significant relationship between the two facts. Column (5) of Table 1.14 and Figure 1.5 show that there is a small and insignificant effect of former peer force use on an officer's force use. Therefore, we can conclude that our results are not driven by officers mimicking the force use of their former peers.

Table 1.14: Effect of Past Peer Force Use on Officer Force Use

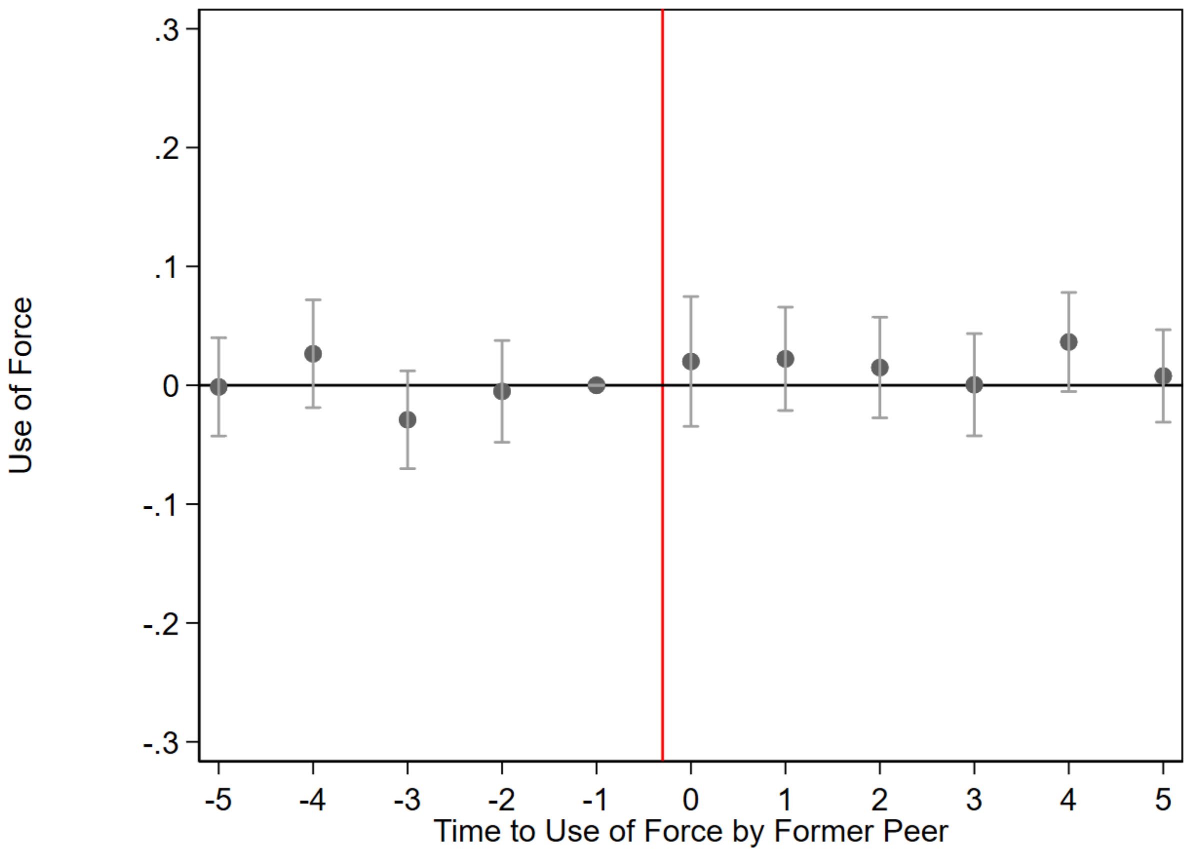
	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force	(6) Force
Former peer in previous week	0.00512***	0.00373***	0.00171***	0.000535	0.000472	0.0288*
	(0.000629)	(0.000606)	(0.000381)	(0.000337)	(0.000318)	(0.0171)
Constant	0.0150***	0.0157***	0.0168***	0.0174***	0.0174***	-3.022***
	(0.000518)	(0.000555)	(0.000439)	(0.000397)	(0.000168)	(0.0103)
Model	OLS	OLS	OLS	OLS	OLS	Poisson
Percent Increase	34.22	24.9	11.45	3.57	3.16	2.88
Unit-Week Fixed Effects	NO	YES	YES	YES	YES	YES
Number of Former Peers	NO	NO	YES	YES	NO	NO
Test Period Fixed Effects	NO	NO	NO	YES	NO	NO
Individual Fixed Effects	NO	NO	NO	NO	YES	YES
Pre-trend Test	0.000	0.000	0.000	0.358	0.355	0.468
R-squared	0.000	0.022	0.023	0.024	0.040	
Observations	986.111	986.088	986.088	986.088	986.088	607.688

Note: Columns 1 through 7 display coefficients from estimates of Equation 1.2 where the outcome variable is an indicator representing whether the officer used force and the event is whether the officer's former peer used force but was not injured in the previous week. We calculate the percent increase by dividing the Column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Figure 1.5 displays all percent changes from Column (5) in this regression testing for differences in pre-trends. Column (5) in Appendix Table 1.22 displays all of the coefficients on lags and leads for Columns (5) and (6) in this table. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

1.6.2 Officers Increasing Effort

Next, we investigate whether officers increase their effort following an injury to a former peer. A potential issue is that officers have to work more hours after a peer injury because the peer has been removed from the pool of eligible workers. This is unlikely for two reasons. First, we are using events that occur to former peers. That means that the injured officers

Figure 1.5: The Effect of Past Peer Force Use on Officer Force



Note: The graph shows coefficient estimates using Equation 1.3 where the outcome is the use of force and the event is the use of force of former peers who did not experience an injury. Coefficients are divided by baseline rate of force and 95% confidence intervals. The baseline rate of force is calculated as the constant term from a regression of force on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer. Red vertical line represents the injury week.

work in a separate police district. Second, there is no reason to expect that peers should be differentially affected by such events.

To investigate potential increases in police effort, we follow (Mas, 2006) and (Ba and Rivera, 2019), we use arrests as a measure of officer effort. Column (1) in Table 1.15 shows the impact of former peer injuries on the probability of arresting a suspect for any reason in the following week. Figure 1.6 displays the dynamics. Overall, we find that there is an insignificant and economically small impact on officer's effort.

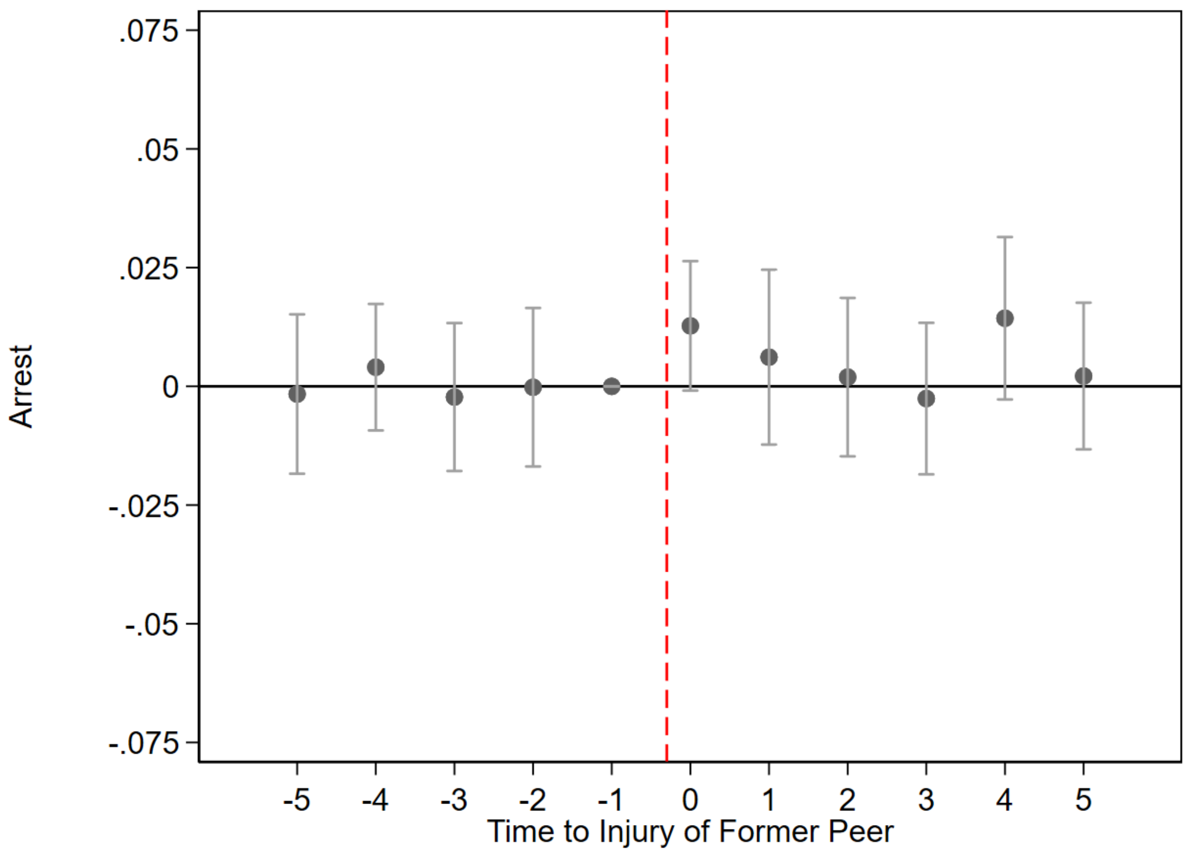
Table 1.15: Effect of Past Peer Injuries on Officer Arrests

	(1) Any Arrest	(2) Non-Index Crime	(3) Property Crime	(4) Violent Crime
Former peer in previous week	0.00249	0.00278*	0.000258	-0.000187
	(0.00202)	(0.00164)	(0.000949)	(0.000941)
Constant	0.299***	0.176***	0.0481***	0.0767***
	(0.000233)	(0.000189)	(0.000109)	(0.000108)
Model	OLS	OLS	OLS	OLS
Percent Increase	0.84	1.6	0.54	-0.25
Unit-Week Fixed Effects	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES
Pre-trend Test	0.871	0.698	0.646	.122
R-squared	0.280	0.254	0.067	0.075
Observations	933,782	933,782	933,782	933,782

Note: Columns 1 through 8 display coefficients from estimates of Equation 1.2 where the outcome variable is an indicator representing arrests for various types of crime. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Appendix Table 1.23 displays all coefficients from this regression testing for differences in pre-trends. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Columns (2) through (5) in Table 1.15 shows the impact of former peer injuries on different types of arrests. The arrest types correspond to the crime types: violent, property, and non-index crimes. We find that non-index crime arrests increase by 1.6% in the week following a peer injury. In contrast, property and violent crime arrests remain unaffected. These results suggest that officers do not materially change their effort after they experience a peer injury.

Figure 1.6: The Effect of Past Peer Injuries on Arrests



Note: The graph shows coefficient estimates using Equation 1.3 divided by baseline rate of arrests and 95% confidence intervals. The baseline rate of arrests is calculated as the constant term from a regression of arrests on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer. Red vertical line represents the injury week.

1.6.3 Social Learning and Emotional Responses

Now that we have ruled out officers increasing effort or mimicking their peer's force use, we seek to determine why officers respond to peer injuries. We consider whether social learning (Banerjee, 1992; Bikhchandani et al., 1992) or emotional responses (Rick and Loewenstein, 2008; Kőszegi, 2006) are more likely to drive the results.

Social learning will drive the results when officers are more likely to learn about their former peers' injuries and use this information in their future decisions to use force. Injuries to former peers may act as a private signal of the underlying injury risk. The signal could cause officers to update their beliefs about the probability that a non-compliant civilian will injure them, increasing their propensity to use force.

On the other hand, previous literature has also documented that negative emotional states can influence an individual's propensity to engage in violence (Card and Dahl, 2011; Munyo and Rossi, 2015; Eren and Mocan, 2018). Several laboratory experiments also show that exposure to violence can affect time and risk preferences. Loewenstein (1996) documented that preferences can be malleable and can be temporarily affected by emotional states. For example, traumatic events and natural disasters can impact risk-preferences (Cameron and Shah, 2015; Tanaka et al., 2010; Hanaoka et al., 2018). Similarly, Hjort (2014) finds that animus discrimination can increase in response to ethnic conflict, and Rohlfs (2010) finds that exposure to violence can make individuals more violent.

Two main predictions separate these mechanisms. First, if social learning is driving the results, officers would have a lower chance of experiencing an injury themselves in the week following a peer injury as they have better information about their true injury risk while on duty. We investigate this by estimating Equations 1.3 and 1.2, with the outcome being an indicator representing the officer's injury status. We report these results in Figure 1.7 and Column (5) of Table 1.16. We find that injury risk falls by 0.0181 percentage points, or 7.26% in the week after a former peer is injured. However, these results are not statistically significant at conventional levels.

Table 1.16: Effect of Past Peer Injuries on Officer Injuries

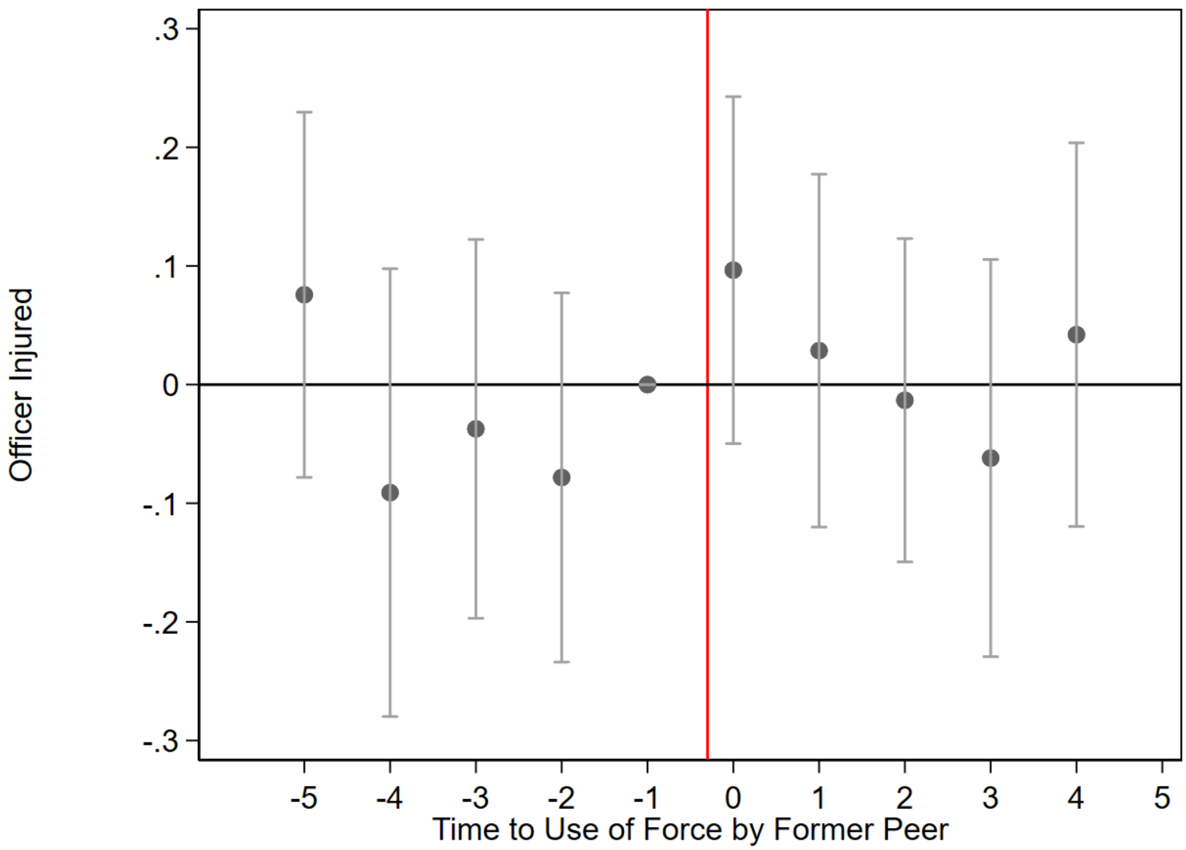
	(1) Injured	(2) Injured	(3) Injured	(4) Injured	(5) Injured	(6) Injured
Former peer in previous week	0.000205	0.0000421	-0.000142	-0.000194	-0.000168	-0.0759
	(0.000189)	(0.000177)	(0.000167)	(0.000166)	(0.000165)	(0.0706)
Constant	0.00231***	0.00233***	0.00235***	0.00236***	0.00236***	-3.102***
	(0.0000710)	(0.0000781)	(0.0000672)	(0.0000610)	(0.0000186)	(0.00859)
Model	OLS	OLS	OLS	OLS	OLS	Poisson
Percent Increase	8.86	1.82	-6.13	-8.37	-7.26	-7.59
Unit-Week Fixed Effects	NO	YES	YES	YES	YES	YES
Number of Former Peers	NO	NO	YES	YES	NO	NO
Test Period Fixed Effects	NO	NO	NO	YES	NO	NO
Individual Fixed Effects	NO	NO	NO	NO	YES	YES
Pre-trend Test	0.000	0.003	0.054	0.108	0.085	0.084
R-squared	0.000	0.020	0.020	0.020	0.026	
Observations	986,111	986,088	986,088	986,088	986,088	88,582

Note: Columns 1 through 7 display coefficients from estimates of Equation 1.2 where the outcome variable is an indicator representing whether the officer experienced an injury. We calculate the percent increase by dividing the column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. Figure ADD displays all percent changes from Column (5) in this regression testing for differences in pre-trends. Column (5) in Appendix Table 1.17 displays all of the coefficients on lags and leads for Column (6) in this table. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Second, the two mechanisms will also deferentially affect the dynamics of the effects. Under social learning, former peers learn about the treatment effects more quickly than those who are not former peers. Then, the treatment group should have a constant effect while the control group's propensity to use force should increase to match the treatment group. In contrast, under an emotional responses mechanism, the effects should dissipate because officers stop responding within a few weeks. Loewenstein and O'Donoghue (2007) point out that the temporal proximity to the event greatly impacts emotional responses. Indeed, Card and Dahl (2011) and Munyo and Rossi (2015) both find that the emotional responses to sports losses are concentrated in a narrow time window after the game.

We display the effects separately by those experiencing (treatment) and not experiencing (control) an injury to a former peer in Figure 1.8. In line with the emotional response mechanism, control officers (those who do not experience a peer injury) do not increase their

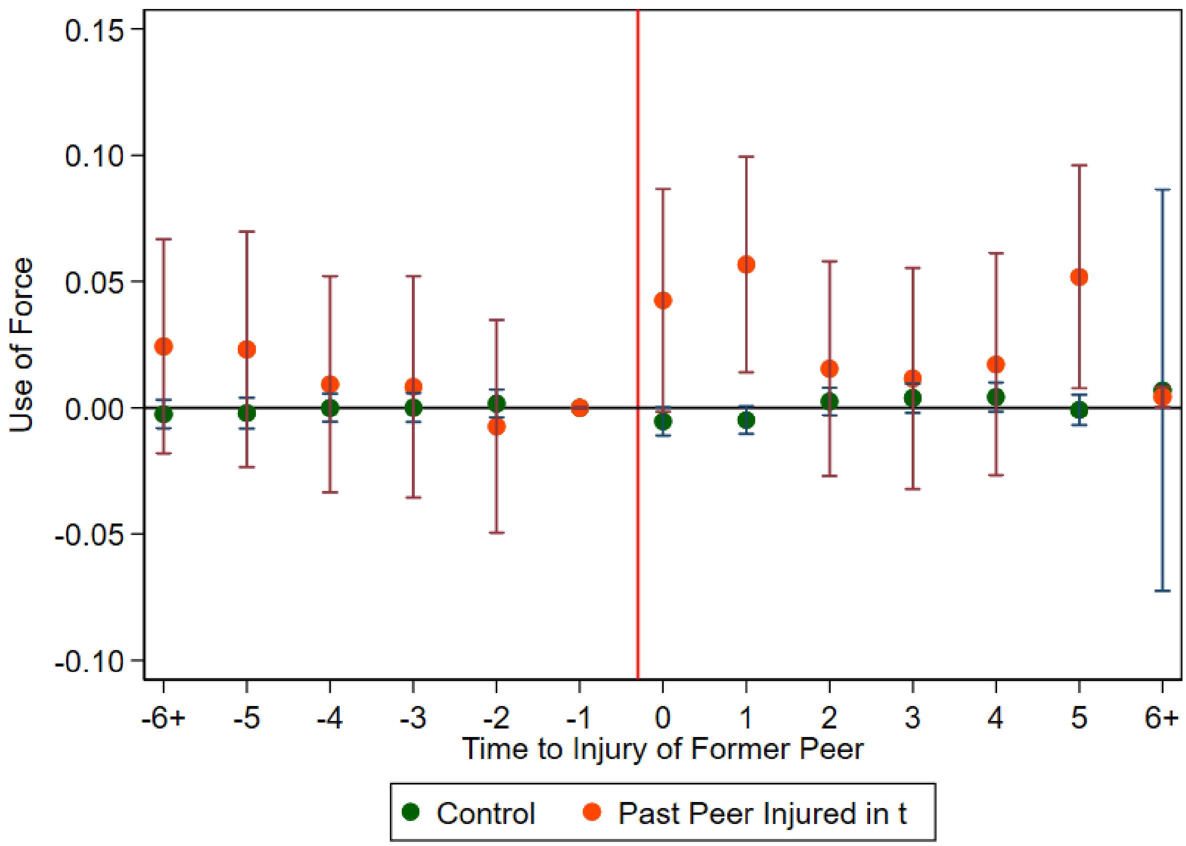
Figure 1.7: The Effect of Past Peer Injuries on Own Injuries



Note: The graph shows coefficient estimates using Equation 1.3 divided by baseline rate of officer injuries and 95% confidence intervals. The baseline rate of officer injuries is calculated as the constant term from a regression of FPS complaints on treatment lags and leads without fixed effects. Standard errors clustered by academy cohort (G = 81). The regression includes individual and unit-week fixed effects. Treatment is an injury of a past peer. Red vertical line represents the injury week.

propensity to use force. In contrast, those experiencing a peer injury (treatment officers) increase their propensity to use force in the weeks immediately after the event and then return to their baseline level of force-use soon afterward. Although we cannot fully rule out a role for social learning, these results strengthen the case for the emotional response interpretation.

Figure 1.8: The Effect of Past Peer Injuries on Use of Force



Note: This figure displays the propensity to use force in periods around officer injuries relative to the week before an officer injury. The control group refers to individuals who did not have a former peer injured. The values are calculated by first regressing force use on individual and unit week fixed effects. The residuals from this regression are then averaged within indicators representing periods of time since an officer injury. We also present 95% confidence intervals using standard errors clustered at the individual level.

1.7 Conclusion

A large and growing body of evidence highlights the harmful effects of policing. Police officers kill roughly 1,000 individuals each year in the United States. Excessive and improper force has also been linked to civil unrest (DiPasquale and Glaeser, 1998), institutional distrust (Bobo and Thompson, 2006; Kirk and Papachristos, 2011), and reductions in the educational and psychological well-being of racial minorities (Ang, 2021).

This article shows that violent interactions with civilians can have spillovers onto how other officers interact with civilians. Force-related injuries to a police academy classmate working in a different district cause an officer to increase their propensity to use force by 7% and propensity to injure a suspect by 10% in the following week. These effects are largest when the injured peer is of the same race.

We do not find evidence that these effects are driven by officers mimicking their peers' force use, changing their effort, or social learning. Rather, we find that the increase in force-use is concentrated in a narrow time window around the event and that the effects fade with officer tenure. Together, these results suggest an emotional response interpretation.

Given that we focus on peers acquired through the police academy and do not consider effects from peers in the same unit, these effects likely underestimate the total effect of peer injuries. The existence of spillovers in police force resulting from on-the-job injuries has important implications for policies meant to reduce improper use of force. Policies that increase the risk officers face will also increase the use of force through peer effect externalities. Policies that reduce the risk officers face will increase have the added benefit of reducing instances of force use. Policymakers should take these externalities into account when determining the optimal way to reduce improper use of force. Interventions that increase the risk of injury to officers will have a muted effect on force use when peers respond by increasing force. Alternatively, policies that reduce the risk to officers will have the added benefit of decreasing the amount of force used by officers. The results also suggest that providing counseling or other support services after an officer experiences trauma may reduce the instances of force. Moreover, future research should attempt to identify interventions that can help prevent officer injuries without increasing the risk suspects face.

1.8 Appendix

1.8.1 Construction of Data Set

To construct our treatment and outcome variables, we link administrative unit assignments from the Chicago Police Department to (i) tactical response reports created after a police officer uses force, (ii) arrest data generated after an officer arrests a suspect, and (iii) formal complaints against an officer.

This section describes the linking process and illustrates how we go from the original sample of officers, arrests, instances of force-use, and complaints to the sample we use for analysis.

1.8.2 Police Academy Cohorts

The administrative district assignments data ranges from January 2002 to December 2016. This data set contains the unit assignment of each office who served at any point during this period. Officers in the Recruit Training Unit (Unit 44) are part of the police academy for the first six months in that unit and on a probationary period during the following twelve months with some variation. We construct the final sample if 3,491 officers from the full sample of 29,894 officers by doing the following:

1. We drop 24,533 officers who graduated from the probationary period before January 2002 (24,533) because their cohort cannot be determined.

2. We impute the start month for 196 officers who graduate from the police academy within a year of January 2002 but begin the police academy before January 2002.

3. We drop 973 officers who never leave the Recruit Training Unit during the sample period. The dropped officers include includes 645 officers who begin at the academy in May 2015 or later and 328 officers who start before May 2015.

4. We drop 23 officers who start in the same month with three or fewer other officers because we believe these to be errors.

1.8.3 Unit Assignments

After restricting the data to 3,491 officers, we then match officers to police districts using a monthly unit-assignment panel obtained from the Chicago Police Department. These assignments tell us the police district assignment of each officer throughout the sample period.

The geographic unit assignments begin roughly eighteen months after a police officer enrolls at the academy. We throw out any months where an officer works in a unit that does not have geographic boundaries. These units include the SWAT team, bomb squad, canine units, detectives, ect. Out of the 3,491 officers in the data set, 3,468 spend at least one month in a geographic unit, and ninety-one percent of days are spent in geographic units.

1.8.4 Tactical Response Reports

The primary source of data comes from the Chicago Police Department's Tactical Response Reports (TRR). The CPD requires that officers fill out a TRR after instances of force-use under circumstances that appear to the CPD's use-of-force model. We use data from TRRs filed between January 8th, 2004 to October 31st, 2016. For every week in the data set, we use this data to measure whether officers use any force in a given week. We also use this data to measure the highest level of force the officers choose to use in that week, if any. Finally, this data is used to measure whether officers or suspects sustain any injuries during their encounters.

A TRR must be filed after use-of-force incidents involving subjects classified as active resisters or assailants. However, some exceptions apply when actively resisting suspects are fleeing, and the members are restricted to verbal commands and/or control holds in conjunction with handcuffing and searching techniques that do not result in the allegation of an injury. For subjects classified as cooperative or passive resisters, police must fill out a TRR if the subject is injured or alleges an injury. A TRR must also be filed for all incidents where a subject obstructs a police officer (Chicago Police Department 2016).

All TRR's require a supervisor's approval. The supervisor must notify the external oversight agency for incidents involving the use of deadly force or the discharge of a firearm, Taser, pepper spray, or other chemical weapons. An external oversight agency must also be notified after an allegation of excessive force.

The variables in the dataset include the date of the incident, number of involved officers, injured officers, suspects' race and ethnicity, injured suspects, and the type of force used against the suspect. One limitation of our dataset is that it includes no records for incidents involving juvenile suspects.

We classify use-of-force incidents into six categories according to the highest type of force used in the incident. Our type-offorce hierarchy comes from the CPD use-of-force model. No and minor force are the only types of force that are authorized for compliant or passively resistant subjects. As mentioned above, TRR's are not required for such incidents unless the subject is injured. We suspect that many incidents involving minor force or less that do not result in injuries are unreported.

Police report injuries to police officers or suspects; the TRR asks explicitly whether the police officer injured the subject. The observed injury rates may reflect some combination of reporting requirements and voluntary reporting.

1.8.5 Arrest Data

Next, we merge onto this data set the arrests made by every officer during this sample period. For every officer, this data set includes every arrest that the officer makes of suspects who are 18 years of age or older. The City of Chicago prevents the disclosure of any information on the arrests of juvenile suspects. These suspects are, therefore, excluded from the analysis.

We restrict the sample of arrests to the same period as the tactical response reports (January 8th, 2004 to October 31st, 2016). Arrest dates are only available from 2010 to 2017. For all years in this study before 2010, we impute the arrest date as the earlier of the bond and release date. Between 2010 and 2017, the median number of days between arrest date and the earliest of the bond and release date is one day (the average is 0.71 days).

For all of the sample's arrests, the arrest data contains a crime code, which describes the reason for the arrest. These codes can designate an arrest for a violent crime, property crime, traffic violation, outstanding warrant, drug crime, municipal code violation, or other violation.

1.8.6 Complaint Data

The complaint data contains all recorded allegations of misconduct filed against officers from 2001 to 2016. The allegations can come from either another officer or a civilian. Each complaint contains information on the officer, complainant demographics, and the date of the incident.

We merge this data to the unit assignment data to measure whether an officer received a complaint about any incident during a given week. We are also able to measure the nature of the complaint. For more information about the complaint data, see Ba (2017).

1.8.7 Background on Lottery

Becoming a Chicago Police officer is a highly sought after career, with thousands of applicants taking the initial entrance exam, which are offered every few years over the past two decades (see Table 1.1). The practice of determining which applicants may enter the police academy based on random lottery number began in the early 1990's as a part of Mayor Daley's attempt to meet racial hiring quotas. This proposal was met with significant uproar and criticism at the time. From one Chicago Tribune article at the time: "Daley came under fire again because new police recruits are being chosen by lottery, not by their performances on the department's entrance exam...The computer then blindly arranged [qualified candidates] in a hiring order that had nothing to do with test results, the officials said" Blau and Kass (2018). From this and other available information, the process is straightforward:

1. Applicants take the test.

2. Passing applicants are given a lottery number randomly generated by a computer.

3. Passing applicants, who are eligible to join the academy, are permitted to enter the academy in order after passing psychological, medical, and physical examinations.

The random lottery process is now accepted by applicants and a common feature of CPD's FAQ's on applying to the department, as FAQs 2018 state: "All applicants who pass the exam are placed on an eligibility list, and that list is sorted in lottery order. You will be referred to the Chicago Police Department in lottery order as vacancies become available" (CPD, 2018). The City of Chicago also uses lottery numbers for training to become a firefighter and EMT (CFD, 2014).

This practice is noted in multiple news articles (Pritchard, 2013; NBC, 2013) and by the Chicago Inspector General (OIG, 2016). While the exact conditions for being drawn in have changed in recent years (after the 2013 test, 21 year-old's were eligible, and preference considerations were made for certain groups such as veterans), two features have remained constant for almost 30 years: lottery ordering and significantly more eligible applicants than spots in the CPD.

Unfortunately, which recruits belong to which cohorts is not able to be obtained through the FOIA to the CPD. A request made in August of 2020 for: "A file containing the date of the test which each officer appointed between 2000 and 2020 took... A file containing the date at which each entrance exam's eligible officer list was retired." was not fulfilled due to excessive burden and noted that "... the Chicago Police Department simply may not compile or maintain in entirety or with the level of detail or sub-categorization you seek..." (FOIA P589445). Based on all available documentation and data, there is no reason to believe a list of eligible applicants is retired when a new test is issued. Rather, it appears to take many months, if not a year, for the first applicants to have their numbers called following a test. This means identifying which cohorts belong to which test for the breadth of our data is not possible.

Summary statistics for these entrance lotteries appears in table 1.1. On average, 85% of test takers pass the entrance exam and 20% of these enter the police academy.[27] We evaluate the balance of the lotteries by performing a multinomial logisitic regression of start month group on the police officers' age, race, and sex. We then use a chi-squared test to determine whether any of the characteristics can predict entrance to a certain police academy cohort. There appears to be some imbalance in two of the nine test-cohorts. This imbalance would be concerning if we were explicitly looking at the effect of contextual effects in police force. However, since the empirical strategy uses a difference-in-differences design the imbalance in these two cohorts will not bias the treatment estimates.

For this reason, we restrict the sample to officers who enter one of 25 geographic districts after graduating from their probationary period. This means that we drop non-standard units such as the canine unit or S.W.A.T. team, who move between geographic districts from day to day. We also drop officers who leave the police academy before six months or individuals who never are registered as leaving the police academy in our sample. We cannot link these data to academy cohorts or the use of force data and cannot be used in the analysis. We also drop thirty-three individuals who have cohort start dates with five or fewer people.

27. There is substantial heterogeneity in the portion of eligible people who enter the academy, ranging from three percent in 2013 to 64% in the first 2006 exam.

1.8.8 Waiting List

Academy cohorts being constructed through a waiting list may raise concerns over identification of treatment effects discussed in (de Chaisemartin and Behaghel, 2020). However, this research discusses the issues associated with treatment being assigned based on randomized waiting lists, where demand for treatment is oversubscribed and treatment effects are estimated by comparing those who received treatment to those who did not. While the CPD academy assignment process is based on lottery numbering waitlists, with far more eligible applicants than spots available, the treatment effect analogous to those discussed in (de Chaisemartin and Behaghel, 2020) is the effect of becoming a Chicago Police Officer (i.e. comparing economic or social outcomes of applicants who entered the academy and those who did not).

In this paper, the population of interest is Chicago Police Officers and the random assignment of police academies is a known peer-group with whom injured officers have social ties but do not experience the same local crime shocks. All of our results are conditional on one being a CPD officer and our control group is not applicants to the department who never had their number drawn. While applying (de Chaisemartin and Behaghel, 2020) to a study of the effect of becoming a police officer would be appropriate, it is not applicable in our environment. Furthermore, the CPD does not provide any information on applicants who did not enter the academy, actual lottery numbers, or waitlists, so constructing a corrected estimator would not be possible.

1.8.9 *Supplementary Evidence*

Table 1.17: Main results Poisson Specification

	(1) Force	(2) Injure Suspect	(3) Arrests	(4) Complaints	(5) Officer Injured	(6) Force
Former peer injured in $t+6$ or earlier	0.0110	0.0260	0.00681	-0.0463	0.0821	0.0302
	(0.0224)	(0.0421)	(0.00463)	(0.0283)	(0.0586)	(0.0322)
Former peer injured in $t+5$	0.00665	0.0505	-0.0000683	0.0276	0.157***	0.00946
	(0.0244)	(0.0472)	(0.00622)	(0.0255)	(0.0604)	(0.0288)
Former peer injured in $t+4$	-0.00508	-0.0933**	0.00274	-0.00408	-0.0804	0.0250
	(0.0194)	(0.0452)	(0.00454)	(0.0227)	(0.0508)	(0.0269)
Former peer injured in $t+3$	0.00337	0.00420	-0.000907	-0.00868	-0.0444	-0.00769
	(0.0244)	(0.0465)	(0.00524)	(0.0212)	(0.0792)	(0.0307)
Former peer injured in $t+2$	-0.0223	-0.0277	0.00150	0.0369	0.0842	-0.0210
	(0.0246)	(0.0412)	(0.00542)	(0.0300)	(0.0835)	(0.0346)
Former peer injured in $t+1$	–	–	–	–	–	–
Former peer injured in t	0.0279	0.0789**	0.00987**	0.00270	0.0409	0.0122
	(0.0225)	(0.0382)	(0.00474)	(0.0313)	(0.0758)	(0.0347)
Former peer injured in $t-1$	0.0531**	0.0906**	0.00506	0.0513**	-0.0710	0.112***
	(0.0263)	(0.0461)	(0.00600)	(0.0254)	(0.0719)	(0.0362)
Former peer injured in $t-2$	0.00823	0.0642	0.00178	-0.0296	-0.0114	0.0497
	(0.0187)	(0.0397)	(0.00527)	(0.0299)	(0.0736)	(0.0318)
Former peer injured in $t-3$	0.0000235	0.0283	-0.000383	0.0418	0.00414	-0.0215
	(0.0240)	(0.0443)	(0.00515)	(0.0317)	(0.0803)	(0.0245)
Former peer injured in $t-4$	0.0106	-0.0253	0.00992*	0.0486*	-0.0766	0.00685
	(0.0265)	(0.0425)	(0.00509)	(0.0276)	(0.0613)	(0.0308)
Former peer injured in $t-5$	0.0425	0.0544	0.00217	0.0548**	0.143***	0.00642
	(0.0262)	(0.0548)	(0.00475)	(0.0262)	(0.0543)	(0.0330)
Former peer injured in $t-6$ or later	0.107**	0.259***	0.200***	0.114*	0.284**	0.121***
	(0.0515)	(0.0822)	(0.0391)	(0.0613)	(0.126)	(0.0369)
Constant	-3.118***	-3.459***	-0.975***	-3.233***	-3.421***	-3.115***
	(0.0488)	(0.0790)	(0.0385)	(0.0602)	(0.123)	(0.0320)
Model	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Peer Definition	Former Peer	Former Peer	Former Peer	Former Peer	Former Peer	Same-Race Former Peer
Unit-Week Fixed Effects	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES	YES	YES
Pre-trend Test	0.840	0.143	0.96	0.593	0.072	0.749
Observations	576,943	218,500	723,630	463,227	825,45	576,943

Note: Columns 1 through 7 display coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing whether the officer used a specific type of force. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.18: Heterogeneous Effects by Type of Force

	(1) Control	(2) No Weapon	(3) Non-Lethal	(4) Baton	(5) Taser	(6) Firearm	(7) Other
Former peer injured in $t+6$ or earlier	-0.000315	0.000538	0.0000821	0.0000349	0.000114	-0.0000389	-0.0000860
	(0.000340)	(0.000414)	(0.0000949)	(0.0000772)	(0.000141)	(0.0000584)	(0.000130)
Former peer injured in $t+5$	0.000533	-0.0000164	-0.00000433	0.0000878	-0.000116	0.0000537	0.000207
	(0.000426)	(0.000459)	(0.0000873)	(0.0000693)	(0.000115)	(0.0000791)	(0.000133)
Former peer injured in $t+4$	0.000232	0.000225	0.000142	0.0000364	-0.0000464	-0.00000922	-0.000109
	(0.000367)	(0.000315)	(0.0000952)	(0.0000925)	(0.000133)	(0.0000562)	(0.000130)
Former peer injured in $t+3$	0.000235	-0.0000989	0.0000507	0.0000802	0.0000231	-0.0000183	0.0000515
	(0.000458)	(0.000434)	(0.0000741)	(0.0000821)	(0.000160)	(0.0000715)	(0.000101)
Former peer injured in $t+2$	-0.0000634	-0.000273	-0.000214**	-0.0000262	0.000166	0.0000134	0.0000463
	(0.000348)	(0.000401)	(0.0000856)	(0.0000734)	(0.000126)	(0.0000616)	(0.000143)
Former peer injured in $t+1$	–	–	–	–	–	–	–
Former peer injured in t	0.000362	0.000665	-0.000101	0.0000985	0.000127	-0.0000511	0.0000839
	(0.000394)	(0.000504)	(0.000100)	(0.0000796)	(0.000140)	(0.0000686)	(0.000102)
Former peer injured in $t-1$	0.000699*	0.000925*	-0.0000229	-0.0000346	0.0000405	0.000118*	0.0000832
	(0.000382)	(0.000498)	(0.0000827)	(0.0000685)	(0.000139)	(0.0000681)	(0.000152)
Former peer injured in $t-2$	0.0000782	0.000101	-0.0000325	0.0000809	-0.000336***	0.0000434	-0.0000741
	(0.000341)	(0.000371)	(0.000109)	(0.0000715)	(0.000122)	(0.0000615)	(0.000139)
Former peer injured in $t-3$	0.000101	0.000319	0.0000331	-0.000129*	0.000000514	-0.0000653	0.0000519
	(0.000340)	(0.000422)	(0.000123)	(0.0000770)	(0.000125)	(0.0000448)	(0.000141)
Former peer injured in $t-4$	0.000135	0.0000922	0.000158*	0.0000224	0.0000544	-0.000126***	-0.00000883
	(0.000381)	(0.000435)	(0.0000943)	(0.0000713)	(0.000151)	(0.0000401)	(0.000107)
Former peer injured in $t-5$	0.000714	0.000882	0.0000587	0.0000113	0.000300	0.0000329	0.00000370
	(0.000438)	(0.000541)	(0.000102)	(0.0000894)	(0.000187)	(0.0000706)	(0.0000943)
Former peer injured in $t-6$ or later	-0.000412	0.00126	0.000337	-0.000223	0.000428	0.0000294	-0.000296
	(0.000978)	(0.00104)	(0.000236)	(0.000204)	(0.000280)	(0.000143)	(0.000287)
Constant	0.0105***	0.0129***	0.000537**	0.000682***	0.00126***	0.000264*	0.00134***
	(0.000941)	(0.000998)	(0.000224)	(0.000192)	(0.000276)	(0.000139)	(0.000267)
Unit-Week Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Pre-trend Test	0.684	0.742	0.010	0.407	0.490 5	0.923	0.545
R-squared	0.035	0.039	0.030	0.022	0.025	0.022	0.023
Observations	944,356	944,356	944,356	944,356	944,356	944,356	944,356

Note: Columns 1 through 7 display coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing whether the officer used a specific type of force. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.19: Effect of Low-Resistance Injuries to Former Peers on Force Use

	(1) Force	(2) Force
Former peer injured in $t+6$ or earlier	-0.000361	-0.0288
	(0.000604)	(0.0308)
Former peer injured in $t+5$	0.0000342	-0.00153
	(0.000749)	(0.0384)
Former peer injured in $t+4$	0.000358	0.0230
	(0.000575)	(0.0281)
Former peer injured in $t+3$	-0.000855	-0.0416
	(0.000708)	(0.0358)
Former peer injured in $t+2$	0.00108*	0.0501*
	(0.000613)	(0.0276)
Former peer injured in $t+1$	–	–
Former peer injured in t	0.000395	0.0220
	(0.000663)	(0.0329)
Former peer injured in $t-1$	0.000339	0.0188
	(0.000537)	(0.0275)
Former peer injured in $t-2$	0.000186	0.0141
	(0.000585)	(0.0276)
Former peer injured in $t-3$	0.0000263	0.00655
	(0.000572)	(0.0283)
Former peer injured in $t-4$	0.000623	0.0310
	(0.000526)	(0.0241)
Former peer injured in $t-5$	-0.000134	-0.00265
	(0.000900)	(0.0443)
Former peer injured in $t-6$ or later	0.00133	0.126***
	(0.000965)	(0.0458)
Constant	0.0164***	-3.120***
	(0.000877)	(0.0424)
Unit-Week Fixed Effects	YES	YES
Individual Fixed Effects	YES	YES
Pre-trend Test	0.113	0.159
R-squared	0.041	
Observations	944,356	576,943

Note: Column 1 displays coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing whether the officer used force and the event is whether the officer's former peer used force but was not injured in the previous week. We calculate the percent increase by dividing the Column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.20: Effect of Past Peer Injuries on Complaints Against Officers

	(1) All Complaints	(2) Force and Verbal	(3) Arrest and Search	(4) Failure to Provide Service	(5) Unbecoming Conduct
Former peer injured in $t+6$ or earlier	-0.000694*	-0.000239	-0.000608**	0.000317	0.00000179
	(0.000413)	(0.000237)	(0.000240)	(0.000191)	(0.0000505)
Former peer injured in $t+5$	0.000518	0.0000688	0.000259	0.000160	0.00000447
	(0.000384)	(0.000251)	(0.000205)	(0.000201)	(0.0000573)
Former peer injured in $t+4$	-0.0000338	-0.000127	0.000211	0.00000106	-0.0000440
	(0.000337)	(0.000182)	(0.000273)	(0.000144)	(0.0000485)
Former peer injured in $t+3$	-0.0000936	0.0000261	0.0000509	-0.000163	0.0000474
	(0.000334)	(0.000220)	(0.000283)	(0.000159)	(0.0000547)
Former peer injured in $t+2$	0.000586	0.000219	0.000151	0.0000953	-0.0000737
	(0.000456)	(0.000192)	(0.000335)	(0.000144)	(0.0000458)
Former peer injured in $t+1$	–	–	–	–	–
Former peer injured in t	0.000113	0.000192	-0.000120	-0.0000603	-0.0000193
	(0.000471)	(0.000213)	(0.000286)	(0.000167)	(0.0000472)
Former peer injured in $t-1$	0.000818**	-0.000165	0.000440**	0.000379**	0.0000165
	(0.000382)	(0.000200)	(0.000204)	(0.000179)	(0.0000595)
Former peer injured in $t-2$	-0.000374	-0.000212	-0.000256	-0.000228	0.000146**
	(0.000472)	(0.000222)	(0.000219)	(0.000167)	(0.0000566)
Former peer injured in $t-3$	0.000800	0.000339	0.000397**	0.0000912	-0.0000437
	(0.000489)	(0.000242)	(0.000192)	(0.000170)	(0.0000497)
Former peer injured in $t-4$	0.000758*	0.000338*	0.000308	0.0000245	-0.0000406
	(0.000417)	(0.000174)	(0.000233)	(0.000149)	(0.0000601)
Former peer injured in $t-5$	0.000832**	0.000265*	0.000202	0.000122	0.00000685
	(0.000377)	(0.000157)	(0.000296)	(0.000180)	(0.0000590)
Former peer injured in $t-6$ or later	0.00151	0.000711	0.00132***	-0.000756	-0.000124
	(0.000963)	(0.000474)	(0.000489)	(0.000535)	(0.000149)
Constant	0.0116***	0.00265***	0.00374***	0.00321***	0.000383**
	(0.000954)	(0.000443)	(0.000493)	(0.000502)	(0.000149)
Unit-Week Fixed Effects	YES	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES	YES
Pre-trend Test	0.412	0.516	0.117	0.468	0.550
R-squared	0.040	0.034	0.038	0.028	0.027
Observations	944,356	944,356	944,356	944,356	944,356

Note: Columns 1 through 5 display coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing types of complaints against the officer. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.21: Heterogeneous Effects by Suspect Characteristics

	(1) White Suspect	(2) Minority Suspect	(3) Male Suspect	(4) Female Suspect
Former peer injured in $t+6$ or earlier	-0.00000784	0.000414	0.000615	-0.000126
	(0.000120)	(0.000428)	(0.000485)	(0.000192)
Former peer injured in $t+5$	0.0000639	0.000326	0.000308	0.0000276
	(0.000118)	(0.000488)	(0.000474)	(0.000144)
Former peer injured in $t+4$	0.0000576	-0.000101	0.0000161	0.0000311
	(0.000122)	(0.000374)	(0.000382)	(0.000175)
Former peer injured in $t+3$	0.0000986	0.000114	0.000156	-0.0000457
	(0.000104)	(0.000474)	(0.000468)	(0.000137)
Former peer injured in $t+2$	-0.0000727	-0.000244	-0.000223	-0.0000971
	(0.0000801)	(0.000469)	(0.000389)	(0.000177)
Former peer injured in $t+1$	–	–	–	–
Former peer injured in t	0.000153	0.000528	0.000799*	-0.000126
	(0.0000933)	(0.000458)	(0.000454)	(0.000146)
Former peer injured in $t-1$	0.0000873	0.00105**	0.00112*	0.0000300
	(0.000120)	(0.000513)	(0.000579)	(0.000145)
Former peer injured in $t-2$	0.0000536	0.000208	0.000385	-0.0000945
	(0.000124)	(0.000337)	(0.000336)	(0.000164)
Former peer injured in $t-3$	-0.000131	0.000376	0.000207	-0.0000110
	(0.0000807)	(0.000446)	(0.000443)	(0.000184)
Former peer injured in $t-4$	-0.000125	0.000196	0.000136	-0.0000139
	(0.000139)	(0.000445)	(0.000437)	(0.000158)
Former peer injured in $t-5$	0.000123	0.000811	0.000840	0.000160
	(0.000139)	(0.000561)	(0.000508)	(0.000139)
Former peer injured in $t-6$ or later	0.0000615	0.000569	0.000442	-0.0000674
	(0.000221)	(0.00120)	(0.00114)	(0.000372)
Constant	0.00103***	0.0154***	0.0143***	0.00261***
	(0.000210)	(0.00113)	(0.00110)	(0.000356)
Unit-Week Fixed Effects	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES
Pre-trend Test	0.626	0.864	0.890	0.984
R-squared	0.036	0.040	0.040	0.026
Observations	944,356	944,356	944,356	944356

Note: Columns 1 through 3 display coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing whether the officer used a specific type of force against a suspect of a given race. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.22: Effect of Past Peer Force Use on Officer Force Use

	(1) Force	(2) Force
Former peer used force in $t+6$ or earlier	0.000271	0.00895
	(0.000330)	(0.0187)
Former peer used force in $t+5$	-0.0000203	-0.00766
	(0.000304)	(0.0159)
Former peer used force in $t+4$	0.000387	0.0183
	(0.000333)	(0.0202)
Former peer used force in $t+3$	-0.000424	-0.0228
	(0.000302)	(0.0163)
Former peer used force in $t+2$	-0.0000751	-0.00342
	(0.000315)	(0.0176)
Former peer used force in $t+1$	–	–
Former peer used force in t	0.000293	0.00855
	(0.000401)	(0.0212)
Former peer used force in $t-1$	0.000325	0.0201
	(0.000319)	(0.0174)
Former peer used force in $t-2$	0.000218	0.0117
	(0.000311)	(0.0175)
Former peer used force in $t-3$	0.00000663	-0.00353
	(0.000316)	(0.0170)
Former peer used force in $t-4$	0.000532*	0.0289*
	(0.000306)	(0.0171)
Former peer used force in $t-5$	0.000115	0.00705
	(0.000286)	(0.0156)
Former peer used force in $t-6$ or later	0.00165	0.127
	(0.00205)	(0.0970)
Constant	0.0153***	-3.161***
	(0.00211)	(0.103)
Unit-Week Fixed Effects	YES	YES
Individual Fixed Effects	YES	YES
Pre-trend Test	0.561	0.621
R-squared	0.041	
Observations	944,356	576,943

Note: Column 1 displays coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing whether the officer used force and the event is whether the officer's former peer used force but was not injured in the previous week. We calculate the percent increase by dividing the Column's coefficient by the baseline in a regression without fixed effects. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.23: Effect of Past Peer Injuries on Officer Arrests

	(1) Any Arrest	(2) Non-Index Crime	(3) Property Crime	(4) Violent Crime
Former peer injured in $t+6$ or earlier	0.00182	0.00142	-0.0000751	0.00153
	(0.00165)	(0.00141)	(0.000715)	(0.00123)
Former peer injured in $t+5$	-0.000357	-0.000185	0.0000202	0.000516
	(0.00185)	(0.00145)	(0.000791)	(0.000770)
Former peer injured in $t+4$	0.000883	-0.00124	0.000721	0.00118
	(0.00147)	(0.00102)	(0.00100)	(0.000836)
Former peer injured in $t+3$	-0.000494	-0.000276	-0.000694	-0.00140
	(0.00172)	(0.00144)	(0.000674)	(0.000955)
Former peer injured in $t+2$	-0.0000422	-0.000546	0.00115	0.000585
	(0.00184)	(0.00157)	(0.000785)	(0.00133)
Former peer injured in $t+1$	–	–	–	–
Former peer injured in t	0.00279*	0.00163	0.00142*	0.000851
	(0.00150)	(0.00133)	(0.000779)	(0.000920)
Former peer injured in $t-1$	0.00135	0.00201	0.000144	-0.000643
	(0.00203)	(0.00171)	(0.000964)	(0.000956)
Former peer injured in $t-2$	0.000424	0.000174	0.000246	0.000312
	(0.00184)	(0.00158)	(0.000884)	(0.000838)
Former peer injured in $t-3$	-0.000564	0.000542	-0.000725	0.0000247
	(0.00176)	(0.00152)	(0.000826)	(0.000844)
Former peer injured in $t-4$	0.00314*	0.00193	0.000324	0.00189*
	(0.00188)	(0.00120)	(0.000951)	(0.00109)
Former peer injured in $t-5$	0.000475	0.000524	0.00229***	0.000214
	(0.00170)	(0.00162)	(0.000737)	(0.00102)
Former peer injured in $t-6$ or later	0.0553***	0.0478***	0.00823**	0.00282
	(0.0116)	(0.00760)	(0.00335)	(0.00542)
Constant	0.247***	0.131***	0.0399***	0.0739***
	(0.0115)	(0.00759)	(0.00332)	(0.00529)
Unit-Week Fixed Effects	YES	YES	YES	YES
Individual Fixed Effects	YES	YES	YES	YES
Pre-trend Test	0.871	0.698	0.646	0.122
R-squared	0.283	0.258	0.068	0.076
Observations	895,525	895,525	895,525	895,525

Note: Columns 1 through 8 display coefficients from estimates of Equation 1.3 where the outcome variable is an indicator representing arrests for various types of crime. We cluster standard errors on the police academy cohort level (G = 81). The pre-trend test row presents the p-value from an F-test for which the null hypothesis is that the coefficients of six lead periods in the event-study specification are simultaneously equal to zero. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

CHAPTER 2

QUANTIFYING REPUGNANCE TOWARD PRICE GOUGING WITH AN INCENTIVIZED REPORTING EXPERIMENT

with Rafael Jiménez-Durán and Eduardo Laguna-Müggenburg[1]

Abstract

Thirty-four states have laws against increasing the price of necessary goods during crises, and individuals take costly actions to report violators of these laws. We provide a theoretical explanation for how these consumer reports reflect a willingness to prevent third-party transactions, called repugnance by (Roth, 2007). We then introduce a new experimental procedure to measure revealed preferences against illicit transactions with imperfect enforcement, called an Incentivized Reporting Experiment (IRE). We implement the experiment during the first wave of COVID-19 to measure the willingness to pay to report sellers who increase personal protective equipment prices. Subjects in the experiment pay to report sellers for price-gouging even when the majority of individuals are willing to buy the goods at the price the sellers charge. The willingness to pay to report is non-negligible, polarized, and responsive to price. Additionally, the mechanism driving the repugnance towards price-gouging is good specific. We find evidence of distaste for firm profits for hand sanitizers, but not face masks.

1. We are grateful to Chris Blattman, Leonardo Bursztyn, Michael Cuna, Steven Durlauf, Julio Elías, Matthew O. Jackson, John List, Nicole Ozminkowski, Marta Prato, Alvin Roth and many more for useful comments. Laguna-Müggenburg acknowledges the support of the B.F. Haley and E.S. Shaw Fellowship for Economics through a grant to the Stanford Institute for Economic Policy Research. This research is funded by the Political Economics Initiative at the Becker Friedman Institute at the University of Chicago. Our study was approved by the University of Chicago Institutional Review Board, IRB20-0448. The experiment was pre-registered at the AEA RCT Registry (#0005597).

2.1 Introduction

Emergencies like natural disasters or pandemics create ideal conditions for prices of essential products to increase. There is typically an increased demand for certain products paired with an inelastic short-run supply or even supply disruptions (Cavallo et al., 2014). The first wave of the COVID-19 pandemic was no exception. There was a surge in demand for personal protective equipment (PPE) exceeding short-run production capacity. Ninety-percent of U.S. mayors reported PPE shortages. One-third of medical facilities urged donations of personal masks to make up for the insufficient supply (Kamerow, 2020). The sharp increase in demand led to dramatic price increases in online marketplaces (Cabral and Xu, 2020). In response, national and state-level emergency declarations triggered price controls on PPE and other essential goods. Thirty-three attorneys general urged companies to help to prevent price-gouging (Selyukh, 2020).

Anti-price gouging laws are ubiquitous; thirty-four states prohibit either increases above pre-crisis prices, 10-20% price increases, or "unconscionable" price increases. Despite their wide adoption, these laws remain controversial among economists. Under perfect competition, keeping prices artificially low can create shortages and causes markets to clear through other margins such as queues or search efforts (Becker, 1965; Barzel, 1974; Weitzman, 1991). However, these policies may improve allocative efficiency under imperfect competition.[2]

An additional factor that complicates welfare evaluations about these laws is that individuals might regard price-gouging as a repugnant transaction, as Roth (2007) argues. Roth (2007) calls a transaction repugnant when when third parties disapprove of and act to prevent it. Individuals can consider a transaction to be repugnant even when they are willing to take part in it (Roth, 2007, 2015). In other words, individuals could get negative utility from others trading essential goods at high prices and want to avoid them.

2. It is well known since at least Pigou (1920) that price controls can restore efficiency with monopolies (Bronfenbrenner, 1947).

This paper proposes a new measure of individuals' repugnance towards price-gouging and provides evidence on its mechanisms. We conduct what we call an Incentivized Reporting Experiment (IRE) in which we measure individuals' willingness to pay to report price gouging. This method takes advantage of the fact that authorities rely on reports from customers to enforce anti-price gouging regulations, as with many crimes (Akerlof and Yellen, 1994). Through the lens of a model, we argue that reporting decisions reflect how much individuals expect to change repugnance with their report and how much they value punishing sellers. Thus, researchers can use this method to measure externalities or repugnance in other settings that also rely on reports to enforce laws or regulations, or even more general settings where consumer actions (e.g., boycotts or negative reviews) impact others.[3] Intuitively, there are two markets: the market for goods, which has an externality (in our case, repugnance), and the "market" for reporting sellers. We argue that the consumer surplus in the reporting "market" contains information about the externality in the goods market.

We operationalize the (pre-registered) framed field experiment as a nationally representative survey distributed by a survey company, CloudResearch. We develop an algorithm that combines text analysis and image recognition to make a list of PPE products (face-masks or hand-sanitizers) listed on Amazon. We randomize subjects into treatments where they make incentive-compatible choices between receiving a gift card and reporting a seller from our list to the Department of Justice. We randomize whether subjects have the option to report a seller who charges either a low ($7.50 - $10) or a high price ($27.50 - $30) amount. Both price ranges represent increases from pre-crisis levels (12-70% and 310-400% , respectively).

We choose the seller at random from the pool of listed sellers and we do not give individuals the seller's information. Hence, reporting decisions reflect only repugnance to price gouging and not other confounders such as the possibility of getting compensation from the

3. For example, Ba (2018) studies the willingness to pay to report police malfeasance in Chicago. Our method offers an alternative to Ba's that does not depend on the existence of naturally occurring exogenous variation in the costs of reporting. We thank Julio Elías for pointing out the similarity with consumer reviews.

seller or reducing own search costs in the future. We use the responses to estimate the subjects' Willingness to Pay to Report (WTPR) sellers.

We use a complementary donation experiment to help tease out the mechanisms underlying the repugnance toward price-gouging. Through the lens of our model, there are two main reasons why individuals might report an unknown seller. Individuals may have a distaste for seller profits or a desire to help others purchase the product at lower prices or with reduced search costs.[4] To tease out between the two, we ask subjects to choose between a $5 gift card and having us donate PPE we purchase from a seller to a hospital. As before, we randomize whether we buy from a high or low-price seller. Since we hold the quantity of PPE donated fixed, donation rates that decrease with higher ask-prices are consistent with a distaste for firm profits, as we argue with our theoretical framework. We also elicit the subjects' willingness to pay for face masks and hand sanitizer, leveraging the fact that our algorithm produces a list of available sellers and that searching is costly.

We provide four sets of results. First, individuals take costly actions to enforce price-ceilings. Eighty-percent of subjects choose to forgo money to report sellers in the lower-price range. On average, the willingness to pay to report sellers who charge the lower-price range was $4.78.[5] Half of the subjects are willing to purchase products in this range, which is an essential component of repugnance.[6] At the same time, seventeen percent of individuals are willing to pay to prevent us from reporting sellers. This polarization is consistent with the findings of Elías et al. (2019) who find that some individuals strongly oppose paid kidney

4. Another reason for reporting a seller would be simply a direct taste for punishing deviations from a social norm of unfairness, as in Kahneman et al. (1986). This does not threaten our interpretation, as long as the taste does not depend on the seller's price. A model in which this taste depends on the seller's price would be hard to be empirically tested against a model in which people care directly about seller profits. For instance, we would need to have the seller burn their profits after charging a high price. Thus, to be more precise, in this paper we group distaste for firm profits with any tastes to punish violations from the social norm of not raising prices in emergencies that vary with the price level.

5. Consider that the price per survey response in these survey companies is around $1.25 for a 10-minute survey.

6. For repugnance to occur, there needs to be individuals and sellers willing to transact.

transactions while others are in favor of them.

Second, the WTPR is increasing in the price that the seller charges, as indicated by our theoretical framework. A one-percent increase in the ask-price increases the WTPR by 0.17%. This increase reflects a shift in the whole willingness to pay distribution. Our finding contrasts with the findings of Elías et al. (2019), where the amount of compensation for kidney transplants did not affect support for these transactions.

Third, while the reporting behavior is similar for both products, the underlying mechanism is likely different. Donation rates decrease by 30% when we buy the PPE from higher-priced sellers, but only for hand-sanitizers; face mask donations are unaffected by the seller's price. Through the lens of our model, this finding suggests that there is distaste for profits generated by hand-sanitizer transactions, but not in masks transactions.

Finally, reporting and donating are positively associated and over 46% of participants are willing to forgo \$5 to have us donate the PPE. Half of the subjects who are willing to pay to report sellers are also willing to forgo the \$5 gift card to have us donate PPE from a price-gouging seller. This result suggests that individuals simultaneously internalize the desire to complete transactions and prevent them from occurring. They are against the transaction when other consumers pay for it but in favor when it is the experimenters who pay for it on behalf of a hospital. Hence, one cannot simply partition the population into those who want to transact and those who find the transaction repugnant.[7] This finding is along the lines of Elías et al. (2019), where support for compensation for kidney donations increases when payments come from a public agency.

Our experiment captures a natural setting. Using observational data from actual price gouging consumer reports filed with different attorney generals, we document that complaints were on the rise during our study period and that the products we chose were prevalent in these complaints. The complaints contain wording that is associated with repugnance, such

7. We thank Al Roth for pointing out this insight.

as "take advantage of people". Moreover, our results are robust to experimenter demand concerns and other confounders such as quality and attention differences.

This paper contributes to three strands of literature. First, we contribute to the literature on repugnance by obtaining a revealed preference measure of repugnance towards price-gouging. Identifying repugnance requires a setting where individuals willingly engage in repugnant transactions and third-parties can restrict the choice set of the potential transactors. This is challenging since many repugnant transactions are prohibited by law. For this reason previous studies primarily use hypothetical vignettes to study repugnance (see Ambuehl et al. (2015) and Elías et al. (2019)).[8] We introduce a revealed-preference method of measuring repugnance, which can be used in other settings that rely on reports for enforcement. Additionally, the only other paper that we are aware of that formally defines and models repugnance is Ambuehl et al. (2015).[9]

Relatedly, Kahneman et al. (1986) argue that community standards of fairness restrict profits attainable by firms; consumers judge firm prices relative to reference levels. Rotemberg (2005, 2011) develops models of consumer anger and firm altruism, where consumers want their sellers to feel altruism towards them. Individuals also judge firms with respect to a reference level. Anderson and Simester (2010) provide experimental evidence of consumer anger along these lines. In our model, the reference level of repugnance is endogenous and depends on the market's distribution of prices.

We also contribute to the literature on price-gouging and price-control regulations. Cavallo et al. (2014) document lower product availability but sticky prices following natural disasters, consistent with a model of "consumer anger" against price increases. In the context of COVID-19, Cabral and Xu (2020) argue that seller reputation might explain why

8. Clemens (2018) uses exogenous variation in migration of guest workers, a job commonly regarded as repugnant, and analyzes the impact of migration of different outcomes (e.g., debt) as loose conditions to test for repugnance.

9. This model, however, relies on an observer misjudging the welfare of a third-party transaction. In contrast, our model does not rely on consumer misperceptions to generate repugnance.

larger and older sellers engage less in price gouging. Chakraborti and Roberts (2020a,b) document increased consumer search following anti-price gouging regulations.[10] Dworczak et al. (2019) develop a model in which the planner does not observe individuals' rates of substitution between a good and money, and find that when there is high dispersion in values for money it might be optimal to impose price controls even if it induces rationing.

Our results suggest that price gouging generates an externality, which a market designer might want to include in welfare calculations of anti-price gouging regulations (Rotemberg, 2008). For example, it might be possible to implement the same allocations in imperfect competition with price controls and subsidies. However, price ceilings might have higher welfare if there is distaste for firm profits since subsidies increase them. Moreover, our results suggest that this depends on the type of product, so a one-size-fits-all policy might not be appropriate in response to emergencies.

Third, we contribute to the literature on fairness and third-party punishment by adding field context to the subject's decisions. A large experimental literature shows that third-parties frequently impose punishments for unfair economic behavior in the laboratory (Fehr and Fischbacher, 2004; Henrich et al., 2006). Some more recent natural field experiments show that altruistic punishment is rare, does not increase with the severity of the violation (Balafoutas and Nikiforakis, 2012; Balafoutas et al., 2016).

In contrast with Balafoutas and Nikiforakis (2012), we find that the vast majority of subjects are willing to punish others. Balafoutas et al. (2016) suggest that their finding in Balafoutas and Nikiforakis (2012) may be due to concerns about counter-punishment. Consistent with this claim, we find that the propensity to punish decreases the cost of punishment. Moreover, our finding that the WTP to report is increasing in the seller's price is consistent with the punishment decision depending on the severity of the harm inflicted by the seller (Carlsmith et al., 2002; Cushman, 2008; Ginther et al., 2016) and suggests that the

10. Beatty et al. (2020) provide similar evidence.

tasks used in Balafoutas and Nikiforakis (2012); Balafoutas et al. (2016) may have suffered from a flat payoff problem (Harrison, 1992).

Finally, our finding that the mechanisms driving reporting differ by good suggests that not all third-party punishment is driven by altruism. Indeed, our model suggests that a distaste for firm profits drives punishment for price-gouging hand-sanitizer. This finding supports self-report evidence that the decision to report in third-party punishment games is due to anger in response to violations of social norms in some cases (Fehr and Fischbacher, 2004; Fehr et al., 2002; Fehr and Gächter, 2002; Herrmann et al., 2008).

Our results also support Pedersen et al. (2018)'s claim that third-party punishment arises because of the interdependency between the punisher's and victim's welfare. A key difference between masks and hand-sanitizer is that masks mainly benefit others while hand sanitizer mainly benefits oneself. The finding that reporting the price-gouging of masks is driven by external benefits while the reporting of price gouging of hand sanitizer is driven by distaste for profits suggests that third-party reporting is partially driven by a desire to realize positive externalities.

The remainder of our paper proceeds as follows. Section 2.2 describes the setting and institutional context. Section 2.3 introduces our theoretical framework. Section 2.4 describes the subjects and experimental design. Section 2.5 describes the empirical results and Section 2.6 argues for their external validity. Section 2.7 concludes.

2.2 Setting

2.2.1 Observational Data Sources

In addition to the data generated by our experiment (which we describe in Section 2.4), we use data from two other sources. First, we use the Rainforest API to obtain information about search results and individual product characteristics from surgical face masks and hand sanitizer listings on Amazon. Each search reviews roughly 10,000 results for face masks and 1,800 for hand sanitizers. We combine an image recognition machine-learning algorithm and text analysis to filter unrelated products from the search results and to convert prices from different presentations to common units (12 fl oz. for sanitizer, 50 pack for masks). According to our algorithm (see Section 2.8.2 for more details), only 6.3% of face mask search results were surgical face masks and 52% of sanitizer search results were hand sanitizer products.[11] Our algorithm, while precise, introduces measurement error relative to selecting products by hand, so the prices that we obtain should be taken with caution.[12]

We also use a database of actual price gouging complaints that consumers filed with Attorney Generals from 6 different states, which we obtained with Freedom-Of-Information-Act (FOIA) requests.[13] Most states required individuals to fill a form that had at least two sections. In the description of the complaint, individuals included information about the seller, product and price. There was also a section that asked individuals for their suggested remedy (e.g., whether they wanted compensation, refund or something else). We machine-read and parsed the text from these two sections and obtained close to 1,900 observations.

2.2.2 Context

The experiment occurred on April 30th and May 1st, three months after the first confirmed COVID-19 case in the United States (Holshue et al., 2020). At this time, the demand for PPE outpaced production capacity. Ninety-percent of U.S. mayors reported PPE shortages and one-third of medical facilities urged donations of personal masks to make up for the

11. Many results in the face mask category were cloth masks, which we distinguish from surgical masks since the medical community has pointed out differences in their effectiveness (MacIntyre et al., 2015). Many results in the hand-sanitizer search were e.g., soaps.

12. Our product classification algorithm has an accuracy of over 0.95. We rely on a large-scale algorithm since we needed to detect sellers that are not easily detectable by manual search (e.g., Cabral and Xu (2020) use a sample of 14-17 hand sanitizers and masks) and we needed results in real time since many products were quickly removed by Amazon and new versions were continuously appearing.

13. Utah, South Carolina, Wisconsin, Idaho, Missouri and Illinois. We filed FOIA requests with every state and the DOJ, but we only received information from these states.

insufficient supply (Kamerow, 2020). The sharp increase in demand led to dramatic price increases. Cabral and Xu (2020) document that, between January and March 2020, mask and sanitizer prices were equal to 2.72 and 1.8 times the 2019 prices, respectively. Within our sample, we observe an average price ratio of 6 for face masks and 5.3 for hand sanitizers, as compared to December 2019 prices (see Table 2.1).[14] Figure 2.1 shows that the price distribution remained stable throughout our sample period, before and after our experiment, and exhibits large dispersion.

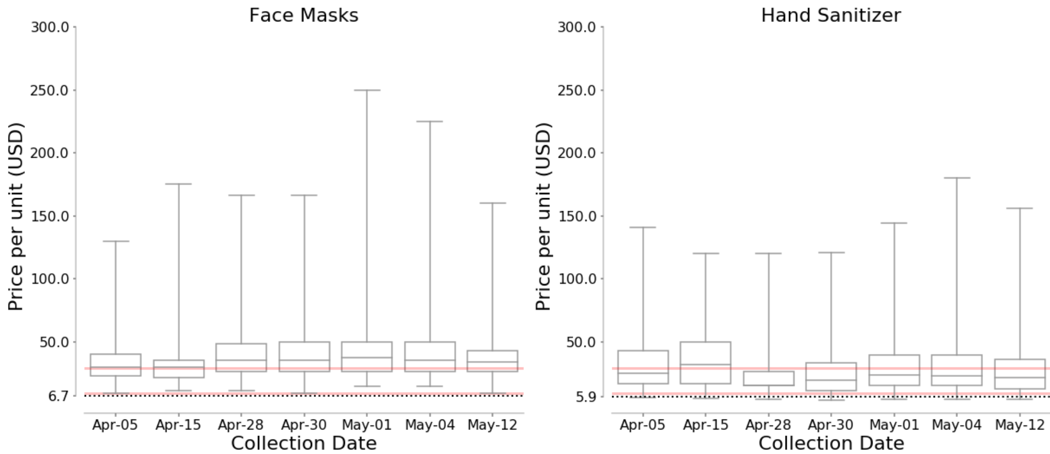
Table 2.1: Personal Protective Equipment Prices in April and May

	Product	N	Price Ratio	Price	Lowest Price	Highest Price
April	Face Masks	1862	5.63	37.74	5.4	349.25
			(4.8)	(32.13)		
	Hand Sanitizer	2251	5.33	31.46	3.49	210
			(4.52)	(26.68)		
May	Face Masks	1122	6.35	42.56	5.99	349.5
			(5.41)	(36.24)		
	Hand Sanitizer	986	5.32	31.38	3.49	220.15
			(5.03)	(29.69)		

Note: Table displays summary statistics for the prices of PPE sold on Amazon between April 5th and May 12th. The prices are normalized to the units of the goods considered in the experiment. The price ratio column displays the average price of the PPE relative to the December price, which was calculated using the data of 4 products obtained from the price-tracking website *camelcamelcamel.com*. This is $6.70 for face masks and $5.90 for hand sanitizer. Standard deviations appear below the means in parentheses. Data scraped from Amazon on April 5th, April 15th, April 28th, April 30th, May 1st, May 4th, and May 12th 2020.

14. The difference between our price ratios and those in Cabral and Xu (2020) could be due to the different sample periods covered; they cover dates between January 15th and March 15th, while we cover April and May. Anecdotally, there was a substantial increase in demand between those dates. Moreover, our sample does not include historical price data; the API only provides real-time data. Our pre-crisis prices come from camelcamelcamel.com and correspond to December prices of 5 sanitizers and 2 face masks that we collected by hand.

Figure 2.1: Amazon Price Distributions on Dates Close to the Experiment



Note: This Figure displays the price distributions on Amazon on dates around our experiment. Boxes contain quartiles of the distributions and the whiskers represent the 1st and 99th percentiles. The pink lines correspond to the price range in our experiment and the dashed lines correspond to the December prices.

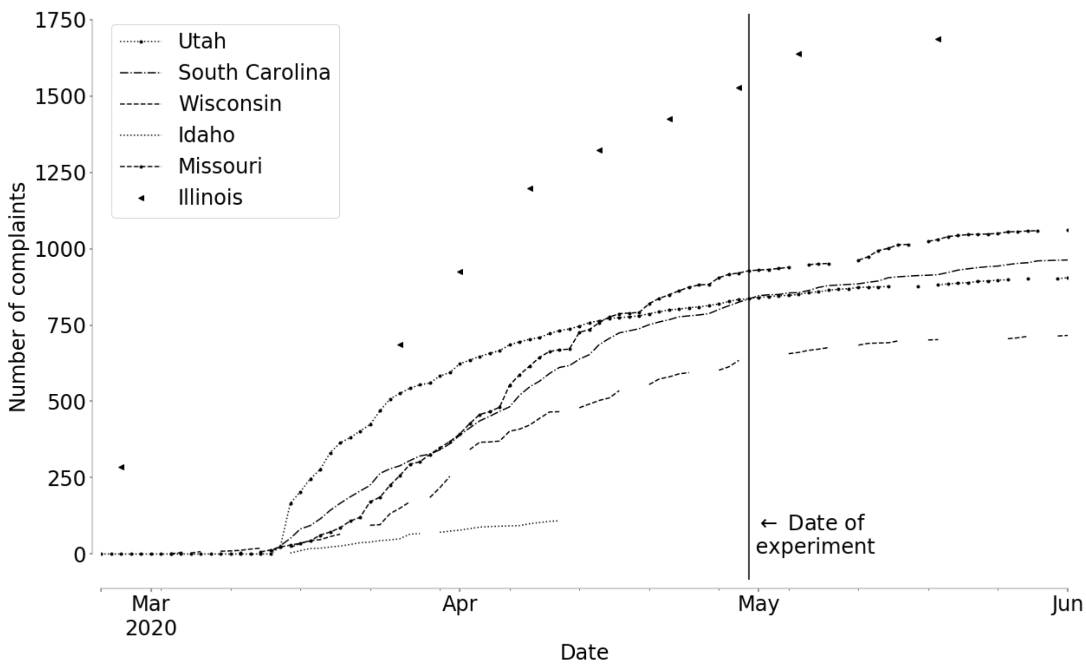
In response to these price increases, Amazon removed over half a million items with excessive prices (Amazon, 2020). State-level emergency declarations triggered price controls on goods "necessary for survival" in thirty-four of these states. These laws prevented either any increases above pre-crisis prices, 10-20% price increases or *unconscionable* price increases. Although there is no federal law against price-gouging, Executive Order 13910 issued on March 23rd prohibited the resale of PPE "at prices in excess of prevailing market prices"

Following the Executive Order, the Department of Justice (DOJ) announced a task force to combat hoarding and price gouging of different products, including sanitizing products and PPE. Individuals could report price-gouging practices to their attorney general or to the Department of Justice's National Center for Disaster Fraud (NCDF).[15] The NCDF requests complainants identify themselves along with the accused, and provide as much information as possible about the transactions. At this point, the complaint is filed and investigated. Individuals found guilty of price gouging face steep fines and up to ten years in prison.

15. See `https://www.justice.gov/disaster-fraud/webform/ncdf-disaster-complaint-form`.

While there is no information about the total number of price gouging complaints received by the DOJ, the states in our sample of complaints had received roughly 1,000 complaints each by the time of our experiment, and they continued to rise afterward. Figure 2.2 plots the evolution of complaints filed in 6 different states. 13% of complaints in our sample include the word "mask" and 10% of them include the word "sanitizer".

Figure 2.2: Cumulative Price Gouging Complaints to States



Note: This Figure displays the cumulative sum of complaints from the beginning of the pandemic to June 2020 for states that responded to FOIA requests.

We summarize the text in our sample of complaints using an unsupervised machine-learning algorithm (latent Dirichlet allocation, LDA) that detects topics automatically from a document.[16] On Table 2.2 we can see that complaint descriptions mostly concern products (e.g., eggs, meat, PPE and toilet paper). On the other hand, consumers refer to "lowering prices", "take advantage of people" and "fair prices" in the section of the forms that asks

16. See Gentzkow et al. (2019) for an overview of LDA topic models and some applications to economics.

about their suggested solution to the complaint. For example, a (selected) complaint filed with the Idaho AG explains that a fair resolution for the complaint is:

"I think they should be fined. I don't want a refund. I want justice."

Table 2.2: Topics from latent Dirichlet Allocation Model

Topic	Prevalence	Top terms
Description		
1	41.4%	egg, dozen, lb, pound, meat, beef, grocery, grind beef, hamburger, dozen egg
2	31.3%	mask, sanitizer, hand, hand sanitizer, bottle, amazon, wipe, lysol, oz bottle, seller
3	27.3%	paper, toilet, toilet paper, gas, station, gas station, towel, charmin, paper towel, gas price
Solution		
1	36.0%	normal, low price, paper, price normal, toilet, toilet paper, difference, desist, bring, cease
2	33.8%	company, gas, raise, complaint, raise price, fix, gas price, report, seller, control
3	30.2%	advantage people, community, accountable, food, hold accountable, fair price, grocery, check, change, issue

Note: The table includes topics from price gouging reports filed to the AGs of Idaho, Illinois, Missouri, and Wisconsin. There are 1890 complaints in our sample (68 from ID, 102 from IL, 1271 from MO, and 449 from WI). "Description" is the field where consumers detail the reason why they are submitting the complaint. "Solution" is the field where consumers express any relief/solution that they are requesting. We only have solutions for 488 complaints. Missouri did not include a field to detail the requested solution. We exclude from the analysis common English stop words and lemmatize the words using the Hunspell dictionary. Top terms are calculated by sorting words according to the $\mathbb{P}[topic|word]$. We decided on three topics for parsimony.

2.3 Theoretical Framework

2.3.1 Setup

We present a simple model to motivate our experimental design and to argue why price gouging complaints contain information about repugnance and about the expected benefits from punishing a seller. The model contains elements from search models and from the volunteer's dilemma of Diekmann (1985). $M > 2$ producers with constant marginal cost $c > 0$ attempt to sell PPE to one of two consumers. Assuming a constant marginal cost is not essential, but it simplifies the exposition.[17] Each consumer has a continuous and weakly decreasing demand for the product, $D(p)$.

First, producers choose prices simultaneously from the equilibrium distribution of prices F over $[\underline{p}, \bar{p}]$. We assume $\underline{p} \geq c$ and take F as given to focus on the equilibrium of the second stage, conditional on F. While interesting, deriving the equilibrium price distribution by incorporating firms' decision-making process is outside of the scope of this paper. Each consumer meets a random producer and observes the ask-price. She then decides whether to accept the offer or keep searching and whether to report the producer for price gouging. Consumers who continue searching match a random unreported producer one more time. Searching costs $c_s \geq 0$ while reporting costs $c_r \in \mathbb{R}$. Reported producers must pay a fine $\kappa \geq 0$. Only one consumer needs to file a report for the seller to be sanctioned and sellers can only be sanctioned once. Reporting is thus a public good; it is useful for removing a producer from the pool of sellers from which people search.[18] Throughout this paper, we

17. Indeed, in the short run marginal costs can be steep. Perhaps this analysis is not valid for the extremely short-run with perfectly inelastic supply, but, as Figure 2.2 shows, price-gouging complaints occur during extended periods of time. In this time frame, it is natural to assume that supply can be adjusted.

18. Intrinsic motivations for reporting a seller can be captured in the cost c_r. Beyond this, this version of the model does not capture additional private motivations for reporting, such as the possibility of a refund if the seller is punished. The model can be adapted without much loss to account for this. However, this matches closely our design, since participants are matched with a random seller and thus are unlikely to be motivated by obtaining a refund.

focus on reports to law-enforcement agencies but the model can be easily adapted to other actions that spillover to other consumers, such as boycotting a product or leaving a negative review online.

Consumers get utility from their own consumption and from the transaction of the other consumer. Given final price offers p_i (paid by the consumer) and p_{-i} (paid by the other consumer), consumer i receives the following payoff:

$$U_i(p_i, p_{-i}) = u(p_i) + s(D(p_{-i})) - \beta\Pi(p_i, p_{-i}) - C_s - C_r.$$

The decreasing function $u(p_i)$ is a monetary measure of welfare derived from direct consumption of the product. In a model without repugnance, this would be the only payoff to the consumer, aside from search C_s and reporting C_r (realized) costs.

In the only other formal model of repugnance that we are aware of (Ambuehl et al., 2015), an observer evaluates the welfare that another individual derives from a given transaction. In our case, the individual derives utility from the other's consumption and disutility from total profits. The function s is weakly increasing to capture positive externalities or social preferences. This form is general enough to account for the other's welfare, consumption or both. Demand D includes consumers' internalization that each unit they consume contributes to firm profits.[19]

The constant β is non-negative. The realized aggregate profits of producers are: $\Pi(p_i, p_{-i}) = D(p_i)(p_i - c) + D(p_{-i})(p_{-i} - c) - K$, where K is the total amount of fees that sellers pay for price-gouging violations. K is equal to κ if consumers report one seller and equal to 2κ if they report two. We assume that revenue $D(p)(p - c)$ has a unique maximum at the monopoly price, p^m, and that for all $p < p^m$ it is strictly increasing.[20]

19. In particular, demand would also be a function of β; $D(p; \beta)$.

20. As we show in Section 2.8.1, the probability of reporting is increasing in price, so if we solved the full equilibrium price distribution, sellers would never choose a price above the monopoly price, hence $\overline{p} < p^m$.

Roth (2015, p. 195) calls a transaction repugnant "if some people want to engage in it and other people don't want them to". Hence, one sensible definition of the repugnance of a given transaction is the (negative) surplus that a third party gets from that transaction: if the surplus is negative—all else equal—the individual would be willing to pay to prevent that transaction from happening. Within the context of our model, the surplus from a third-party transaction originates from the quantity/surplus that the other consumer gets and the profits that the seller gets from that transaction. Our definition our repugnance is the following.

Definition 1. The repugnance that consumer i derives from the transaction between the other consumer $-i$ and the matched seller is given by the surplus:

$$R(p_{-i}) \equiv -\underbrace{\left(\underbrace{s(D(p_{-i}))}_{\text{External benefits}} - \underbrace{\beta D(p_{-i})(p_{-i} - c)}_{\text{Distaste for profits}} \right)}_{\text{Surplus from other's transaction}} \tag{2.1}$$

The goal of this paper is to provide a proxy measure of $R(p)$ and to tease out whether it is mostly driven by distaste for seller profits or external benefits. Note that this repugnance can be positive, in which case the transaction is repugnant in the usual sense of Roth (2007). If $R(p) > 0$, it represents the amount that the consumer would be willing to pay to prevent that transaction from happening. However, if $R(p) < 0$, the transaction would be desirable (negative repugnance), so the consumer would be willing to pay that amount to make that transaction happen.[21]

This setting can capture forms of repugnance beyond price gouging, such as repugnance to kidney transplant markets. For instance, Elías et al. (2019) study repugnance towards kidney transplants. We can also consider preferences for punishing violations of social norms (Fehr and Fischbacher, 2004)—indeed, the equation below shows that payoffs are a function

21. An example of a transaction with negative repugnance would be the movement to "buy local".

of the expected punishment (fees) that the sellers receive if they violate anti-price gouging laws. There are other types of repugnance for which this framework is not ideal, such as eating horse meat: in this case consumers derive negative utility from others' consumption, so the assumption of $s' > 0$ is not reasonable.

We can rewrite the realized payoff of consumer i in terms of repugnance as:

$$U_i(p_i, p_{-i}) = \underbrace{\widetilde{u}(p_i)}_{\text{Net direct welfare}} - \underbrace{R(p_{-i})}_{\text{Repugnance}} + \underbrace{\beta K}_{\text{Seller punishment}} - \underbrace{(C_s + C_r)}_{\text{Search and reporting costs}}$$

where net direct welfare is: $\widetilde{u}(p_i) \equiv u(p_i) - \beta D(p)(p - c)$. This term corresponds to the individual's direct payoffs net of the distaste for firm profits accrued from consumption. Since profits are increasing in price, this term is decreasing.

We are interested in Bayesian Nash equilibria (BNE), conditional on F. Consumers choose behavioral strategies $\sigma^r(z)$ and $\sigma^a(z) \in [0, 1]$, that denote the probabilities of reporting and accepting price offer z, respectively. The expected value of reporting versus not reporting (see Section 2.8.1) is:

$$v^r(z) - v^n(z) = \underbrace{\beta\kappa\left(1 - \frac{\sigma^r(z)}{M}\right)}_{\text{Expected value of punishment}} + \underbrace{\left(\frac{R(z) - \mathbb{E}R}{M}\right)(1 - \mathbb{E}\sigma^a)}_{\text{Expected change in repugnance}} - c_r \tag{2.2}$$

where the operator $\mathbb{E}x$ denotes expected value of x, $\mathbb{E}x = \int_{\underline{p}}^{\overline{p}} x(p)dF(p)$. Equation 2.2 shows that the value of reporting depends on the expected value of punishing the seller, the expected change in repugnance and the cost of reporting. The value of accepting an offer versus searching is:

$$v^a(z) - v^s(z) = c_s + \widetilde{u}(z) - \mathbb{E}\widetilde{u} + \frac{\operatorname{Cov}(\sigma^r, \widetilde{u})}{M} \tag{2.3}$$

where $\operatorname{Cov}(x, y) = \int_{\underline{p}}^{\overline{p}}(x(p) - \mathbb{E}x)(y(p) - \mathbb{E}y)dF(p)$. In Section 2.8.1, we prove the existence of a BNE.

In equilibrium, there exists a reservation price r^* such that consumers accept all offers with price smaller than r^*, and search when they receive more expensive offers. The equilibrium probability of reporting satisfies:

$$\sigma^r(z) = \max\{\min\{\sigma^*(z), 1\}, 0\}$$

where:

$$\sigma^*(z) = \frac{M}{\beta\kappa}\left[\beta\kappa + \left(\frac{R(z) - \mathbb{E}R}{M}\right)(1 - F(r^*)) - c_r\right] \tag{2.4}$$

The probability of reporting is thus (weakly) increasing in the producer's ask-price, z, and decreasing in the cost of reporting, c_r.

2.3.2 Claims

We make four claims regarding the key outcomes of our experiment: reporting and donating.

Claim 1. *There exists a monetary compensation such that an individual is indifferent between reporting and not reporting a seller, when the other individual plays an equilibrium strategy. We will call this compensation the willingness to pay to report (WTPR). This quantity depends on the expected change in repugnance.*

The closed-form expression of the WTPR is:

$$WTPR(z) = \begin{cases} \beta\kappa + \left(\frac{R(z)-\mathbb{E}R}{M}\right)(1 - F(r^*)) - c_r & \text{if } \sigma^*(z) < 0 \\ 0 & \text{if } \sigma^*(z) \in [0, 1] \\ \beta\kappa(1 - 1/M) + \left(\frac{R(z)-\mathbb{E}R}{M}\right)(1 - F(r^*)) - c_r & \text{otherwise} \end{cases}$$

The WTPR depends on the expected value of punishing a seller and on the expected change in repugnance due to the individual report, minus reporting costs. This claim relies heavily on the assumption that a reported seller is removed from the pool of available sellers next period—and thus is unable to match with the other consumer, if she chooses to keep searching. Note that this WTPR might be positive or negative, depending on punishment values, reporting costs, and how far is $R(z)$ from its average. Also, note that the WTPR does not depend on the individual's own consumer surplus, ruling out direct benefits from reporting. This result is due to the fact that reporting does not change the distribution of sellers that the individual can meet in case she decides to search (since there is no way of matching again with the same seller). Our experiment will also rule out any direct benefit from reporting, since we don't give subjects information about the seller other than their price.

Furthermore, the WTPR does not depend on repugnance $R(z)$ directly, but relative to the average "market" repugnance, $\mathbb{E}R$. Hence, we obtain an endogenous reference price that depends on the price distribution and is used to evaluate the repugnance of a transaction for the purposes of reporting it. This is consistent with the literature on fairness, in which community standards of fairness depend on comparisons with respect to a reference price (Kahneman et al., 1986). Indeed, in our sample of actual complaints it is common that consumers request prices to go back to normal (see Table 2.2).

Claim 2. *The willingness to pay to report a seller is increasing in price.*

As the seller's ask-price increases, consumption of the other consumer decreases and firm profits increase, since prices are below the monopoly price. Both of these forces place upward pressure on the subject's value of reporting. There is also an opposite force, since the probability that the other consumer reports is also higher, and thus the incentives to free-ride on the other report are also higher, but this effect does not dominate in a symmetric equilibrium.

In the next two claims, we use the model to motivate experimental variation that can separate different mechanisms for reporting.

Claim 3. *If individuals derive utility from external consumption and demand for the product has positive income effects, individuals will be willing to pay for a donation of the product to the other consumer.*

The claim suggests that we can use variation in a donation experiment to understand whether internal or external consumption concerns drive the decision to report price-gougers. A consumer who gets no utility from the consumption of third-parties or firm profits would always prefer to have money for consumption. However, those with other-regarding preferences may be willing to pay to increase the other's consumption.

Note that demand is a function of the consumer's income I. When we donate an amount $\varepsilon > 0$ of PPE to a hospital (as we detail in the experimental design section below), their utility maximization problem changes to: $\max_{\{x,y\}} u(x + \varepsilon, y)$ s.t. $p_x x + p_y y = I$, where x is the amount of PPE they purchase. This problem is equivalent to solving $\max_{\{x,y\}} u(x, y)$ s.t. $p_x x + p_y y = I + p_x\varepsilon$, so demand of x becomes $D(p; I + p_x\varepsilon)$. If s is increasing and PPE is a normal good, then $s(D(p; I + p_x\varepsilon)) > s(D(p; I))$, so individuals are willing to pay for donations.

Claim 4. *Under the assumptions of Claim 3, if individuals have distaste for firm profits, their willingness to pay for a donation of the product decreases as the price at which the product is purchased increases.*

When we donate a product that we buy from any seller, external consumption is fixed at $D(p; M + p_x\varepsilon)$ and does not depend on the seller's price. When we buy from a more expensive seller who charges $z' > z$, profits increase by $(z' - z)D(p; M + p_x\varepsilon)$.[22] Thus, only subjects with $\beta \neq 0$ should respond to variation in z. Note that one key assumption, embedded in our model, is that supply is not perfectly inelastic, so purchasing from a producer increases

22. We assume that we increase the demand for the products of both sellers so they don't hit their capacity constraints or modify their subsequent pricing decisions. We are also assuming that the subject does not get utility from the experimenter's payoff. This is a standard assumption present in the reporting experiment and experimental elicitations and MPLs.

its profits.[23] Moreover, our model also assumes that individuals are not altruistic towards us; that is, they don't incorporate the experimenter's budget into their welfare. We discuss experimenter demand effects on Section 2.6.

2.4 The Experiment

A survey company, CloudResearch, recruited 1,418 participants from the United States for the experiment. The company selected these participants to match the U.S. census on race, Hispanic origin, age, and gender. The characteristics of our subjects is shown in Table 2.3.

The experiment begins with questions related to purchasing behavior. We then elicit the willingness to pay for PPE and ask subjects to report the lowest PPE price they consider excessive. After the surveys, we assigned subjects into treatments using a 2×2 completely randomized between-subjects design. Treatment is balanced across almost all demographics except education, see Table 2.4. There is some imbalance in education, but controlling for education dummies does not change the coefficients in our regression models, suggesting that this chance imbalance did not affect our results.

The treatments varied the type of PPE subjects would consider independently with a seller's ask-price. Half of the subjects considered a lower price range ($7.50 to $10) and a higher price range ($27.50 to $30). Both price-ranges constitute illegal price-increases under many price-gouging regulations. We induce this variation to test Claims 2 and 4 which rely on comparative statics over the seller's price. Within each price-range we evenly split subjects into treatments that consider 12 FL oz / 355 ML hand sanitizer or 50 count disposable face masks. We use two different types of PPE to investigate good-specific heterogeneity in the willingness to pay to report or the mechanisms. We revealed pre-crisis prices (December 2019) were $5.90 for hand sanitizer and $6.70 for face masks of equivalent presentations,

23. We thank John List for pointing this out to us. If that producer would sell out its stock disregarding whether we buy or not, our treatment would not have any impact on profits.

Table 2.3: U.S. Adult Sample Description

index	Full Sample (N=1439)	Non-attriters (N=1391)	US Pop
Female	52.95	52.91	51.00
Age 18-34	27.94	27.82	32.10
Age 35-54	36.07	36.59	31.30
Age 55+	36.00	35.59	36.60
White (non-Hispanic)	63.72	63.84	62.30
Black	12.09	12.15	12.96
Hispanic	16.54	16.46	16.41
Asian	5.77	5.61	5.96
Other race/ethnicity	2.36	2.44	2.37
Less than HS	2.08	1.94	10.60
HS/GED	15.36	15.31	28.32
Some college/Associate degree	31.97	31.70	27.77
Bachelor's Degree	30.79	30.77	21.28
Graduate Degree	19.81	20.27	12.04
Income $<$ \$50,000	37.53	37.38	43.70
\$50,000 $\leq$ Income $<$ \$100,000	37.46	37.10	30.00
\$100,000 $\leq$ Income	25.02	25.52	26.20

Note: The table describes the demographic characteristics of the respondent sample and compares them to the Vintage 2019 national population estimates from the Census Bureau. The survey company selected participants to match the U.S. census on race, Hispanic origin, age, and gender.The sample over-represents high-education and median income subpopulations based on self-reported information.

to homogenize the points of reference. We also provided a picture of the goods to prevent subjects from confusing disposable face masks with the more expensive N95 face masks.[24]

24. On April 2nd, 2020, Amazon prohibited the sale of N95 face masks on their platform (Rey, 2020).

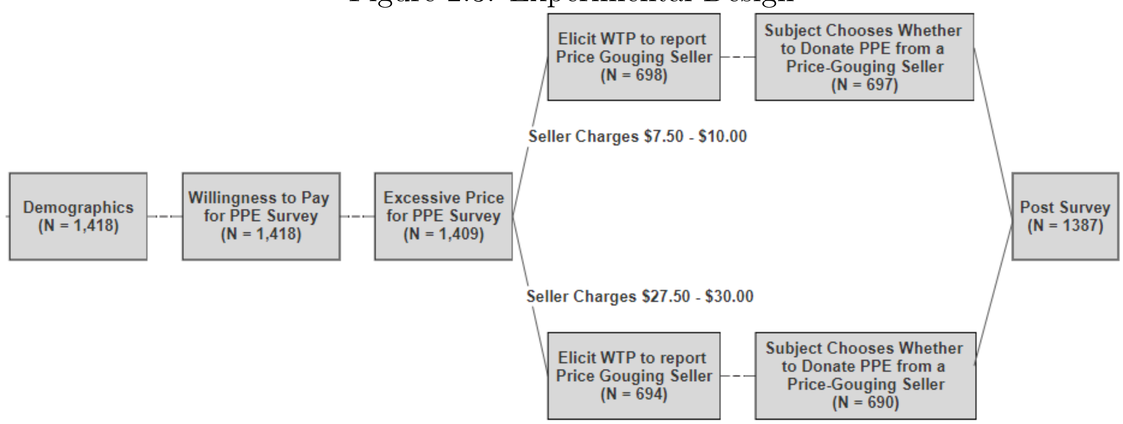
Table 2.4: Treatment Balance

	Hand Sanitizer		Face masks		
	\$7.50-\$10.00	\$27.50-\$30.00	\$7.50-\$10.00	\$27.50-\$30.00	F Test p-value
Age	46.15	45.48	47.32	47.44	
	(17.02)	(16.6)	(17.5)	(16.59)	0.35
Female	0.52	0.55	0.51	0.54	
	(0.5)	(0.5)	(0.5)	(0.5)	0.69
White	0.63	0.66	0.64	0.63	
	(0.48)	(0.47)	(0.48)	(0.48)	0.81
Black	0.12	0.11	0.13	0.12	
	(0.33)	(0.32)	(0.33)	(0.33)	0.95
Hispanic	0.19	0.16	0.16	0.16	
	(0.39)	(0.37)	(0.36)	(0.37)	0.70
Asian	0.05	0.03	0.05	0.07	
	(0.21)	(0.18)	(0.22)	(0.25)	0.21
Other race/ethnicity	0.04	0.05	0.04	0.03	
	(0.19)	(0.22)	(0.2)	(0.18)	0.70
Less than high school	0.02	0.02	0.02	0.01	
	(0.15)	(0.13)	(0.15)	(0.12)	0.79
High school or GED	0.13	0.2	0.18	0.11	
	(0.34)	(0.4)	(0.38)	(0.31)	0.00
Some college/associate degree	0.32	0.34	0.27	0.34	
	(0.47)	(0.48)	(0.44)	(0.47)	0.11
Bachelor's degree	0.34	0.27	0.3	0.33	
	(0.47)	(0.44)	(0.46)	(0.47)	0.16
Graduate degree	0.19	0.18	0.23	0.21	
	(0.39)	(0.38)	(0.42)	(0.41)	0.26
Income $<$ \$50,000	0.38	0.38	0.35	0.38	
	(0.48)	(0.49)	(0.48)	(0.49)	0.83
\$50,000 $\leq$ Income $<$\$100,000	0.46	0.48	0.49	0.45	
	(0.5)	(0.5)	(0.5)	(0.5)	0.69
\$100,000 $\leq$ Income	0.26	0.24	0.25	0.27	
	(0.44)	(0.43)	(0.43)	(0.44)	0.82
Sample Size	349	346	348	348	

Note: Table shows the mean and standard deviations of each variable separately by treatment with standard deviations in parentheses below each mean. The F Test P-value shows the p-value from an F test with the null hypothesis that the means of each variable are equal across all treatments.

Following Kuziemko et al. (2015), we undertook several steps to ensure the sample's validity. First, we only allowed participants with U.S. IP addresses and launched our survey on a workday morning. Second, we included a CAPTCHA to exclude potential robots. Third, we told respondents that payment was contingent on survey completion. Finally, we included attention checks. Figure 2.3 summarizes the flow of the experiment, and an exact copy of the survey appears in Section 2.8.3.

Figure 2.3: Experimental Design



Note: This Figure displays the flow of the experiment. Randomization occurs after the excessive price survey. Within each of the seller price treatments, subjects are randomly split into considering either hand-sanitizer or face masks. Numbers in parentheses represent sample sizes at that stage of the experiment.

2.4.1 Willingness to Pay for Personal Protective Equipment

The survey told subjects about an algorithm we created to track PPE on Amazon. We offered to notify them if the delivery of a similar product was available in two weeks or less. If they wanted to be notified, they could select the maximum price that they were willing to pay for each of the products. At the end of the survey, we provided subjects with a link to a randomly chosen product from our list at or below their maximum willingness to pay. Following our pre-registration plan, we winsorized the data at the 99th percentile.

This procedure is similar to a first-price auction. Subjects have no incentive to quote a maximum that exceeds their valuation of the good—doing so may result in the algorithm showing them a good at a price they are not willing to pay. However, subjects have incentives to report a smaller valuation, trading off higher chances of smaller prices for higher chances of not being informed about a product they are willing to purchase.[25] Throughout the paper we refer to this quantity as willingness to pay for the PPE, with the caveat that it is biased downwards.

2.4.2 Excessive Prices for Personal Protective Equipment

To compare our incentivized measure of willingness to pay to report with hypothetically stated measures about repugnance, we asked subjects to tell us the lowest price they considered to be excessive for both goods. Individuals use numerous adjectives to describe prices in the gouging context, e.g. abusive, unfair, exorbitant or excessive. While all these terms have some normative content and could trigger differentiated concepts in subjects' minds, we chose to use *excessive* as it is commonly used in laws (see, for instance, Giosa (2020)) and describes a situation in which the price is unexpectedly high without placing undue emphasis on potential ill intention of the seller.

2.4.3 Eliciting willingness to pay to report

We elicited the subject's willingness to pay to have us report a randomly chosen seller for price gouging to the Department of Justice using an iterative multiple price list (iMPL). The procedure confronts subjects with an array of paired options and asks them to make a single choice within each pair. At each step, the program asks subjects which of the following two

25. Even if this procedure is not incentive-compatible, it still gives some information about valuations and beliefs about the price distribution. So we opted for it, versus a more contrived, but incentive-compatible exercise. Unfortunately, only 1.2% and 6% of respondents who received a link to a listing of face masks and sanitizer in their price range clicked on it. While this could indicate low consequentiality of our WTP elicitation (few respondents actually willing to "track" products at the price range they report), it could also be due to respondents fearing to lose their progress in the survey (if the Amazon link opens in the same Window). There could be some unobserved fraction of respondents who simply copied the link without clicking on it.

options they prefer:

We **report** an Amazon seller to the **Department of Justice National Center for Disaster Fraud**. This Department is in charge of preventing price gouging for critical supplies. We will report one seller in our list who charges between *[$7.50 - $10.00,$27.50 - $30.00]* for one *[12 FL oz. / 355 ML hand sanitizer, 50 count disposable face masks]*.

You receive a $*[Value]* Amazon Gift card.

All respondents first decide between reporting a seller to the DoJ and a $5 Amazon gift card. If the subject chooses to report, her next decision is between an $8 gift card and reporting the seller. If instead, she selects the money, her next decision is between a $2 gift card and reporting the seller. We continue increasing or decreasing the gift-card amount; Figure 2.4 displays the iMPL's decision tree.

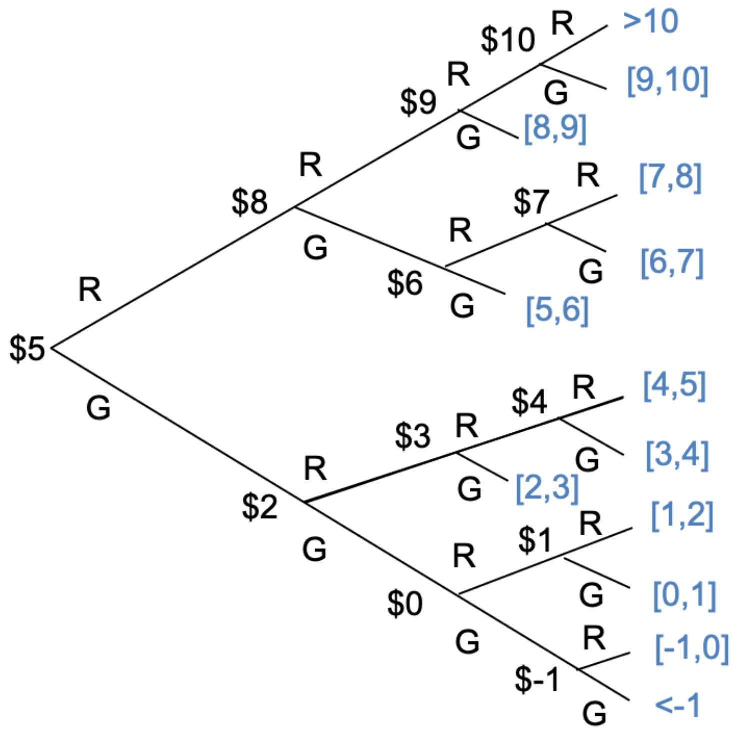
When the differences in values between the last choice and refined choice dropped below $1, the program stopped. We randomly select one in every 10 subjects and randomly implement one of the subjects' decisions (including both reporting and donating).[26]

Variation in the gift card amount maps into variation in c_r and in combination with Equation 2.4, allows us to measure each subject's willingness to pay to report (WTPR). The WTPR can fall into one of thirteen intervals: $(-\infty, -1]$, $(-1, 0]$, $(0, 1]$, $(1, 2]$, ..., $(9, 10]$, and $(10, \infty)$.[27] Following our pre-registration, we either present the portion of subjects falling

26. The iMPL imposes strict monotonicity and enforces transitivity (Gonzalez and Wu, 1999). The method's main advantages are transparency to subjects and avoiding framing effects. However, it provides interval responses rather than an exact WTPR. We elected not to use a method providing exact WTPR's due to concerns of a flat payoff problem (Harrison, 1992). Out of the 1,200 subjects originally planned for the experiment, we selected 120 to implement their choices. From these, 57 were to receive a gift card, 60 to report a seller and 3 to donate a product.

27. To administer the negative WTR, we offered subjects the choice between reporting and earning $1 or not reporting and earning $0.

Figure 2.4: Willingness to Pay to Report Decision Tree



Note: This Figure displays the decision tree subject's faced during the willingness to pay to report the experiment. All subjects began with the decision between a $5.00 gift card and reporting a seller. Subsequent decisions depend on the subject's choice.

within a WTPR interval or set the WTPR value to be the maximum of the interval, 11 in the case of the $(10, \infty)$ interval.

To ensure consequentiality, we chose goods subject to price-gouging legislation. Furthermore, we informed our subjects that our algorithm detected sellers who charged prices between five and fifty dollars in the months before the experiment, so both treatments had the same support. Whenever our algorithm identified sellers charging in a price range at which a subject chose to report, we reported the seller to the NCDF. Thus, report decisions exposed sellers to the threat of steep fines or incarceration.

We do not give participants any information about the seller other than the price. By doing this, we restrict the possibility that they might obtain some direct benefit of reporting, such as reducing their own search costs in the future or obtaining a refund from the seller (as many consumers in our sample of complaints look for).[28] Additionally, this prevents participants from reporting the seller by themselves and still get the gift card, especially since we are saving them the costs of filling out the report form.[29]

2.4.4 Donation Experiment

After the reporting experiment, subjects decided between a \$5 Amazon gift card and Donating PPE to a hospital listed in `getusppe.org`, an organization that allocates PPE donations to health care workers. Under Claim 3, the choices in this stage of the experiment allow us to understand whether individuals derive utility from external consumption. Moreover, we tell subjects that we purchase the PPE from a randomly chosen seller at the price range. The item considered in this step of the experiment matches the iMPL in the type and seller price range. Our treatment thus keeps constant the quantity of PPE donated and varies only the price at which we buy the product. This way, we test our Claim 4 from the theoretical framework. In particular, an individual that receives price offer $z \in \{z_L, z_H\} = \{[\$7.5 - 10], [\$27.50 - 30]\}$ donates if:

$$\mathbb{E}s(D'_{-i}) - \beta z > \mathbb{E}s(D_{-i}) + \$5$$

where D'_{-i} and D_{-i} denote the hospital's demands with and without donation, respectively. Let $\Delta\mathbb{E}s(D_{-i}) \equiv \mathbb{E}s(D'_{-i}) - \mathbb{E}s(D_{-i})$ be the expected change in social preferences. Suppose that each individual in our sample has their own $\Delta\mathbb{E}s(D_{-i})$ and parameter of distaste for profits, β_i. Our test for distaste for firm profits is the difference between the fraction of individuals donating with the low price versus the fraction of individuals donating with the

28. Since there are thousands of search results, the possibility of reducing their own search cost by reporting a random seller is insignificant. However, many other consumers might still match with that seller, so they can still reduce others' search costs, as in our model.

29. Participants could still search for a seller by themselves and report it, but this is true across our treatments and gift-card amounts. Moreover, since there are thousands of noisy search results (see Section 2.2.1), searching is costly

high price. In particular, we hypothesize that, with distaste for firm profits ($\beta_i \geq 0$):

$$\underbrace{\Pr\left(\Delta \mathbb{E} s(D_{-i}) - \beta_i z_L > \$5\right)}_{\text{Fraction donating with low price}} \geq \underbrace{\Pr\left(\Delta \mathbb{E} s(D_{-i}) - \beta_i z_H > \$5\right)}_{\text{Fraction donating with high price}}$$

We presented the question to respondents as

We **buy** from a seller and **donate** to a site listed in `getusppe.org`. This organization coordinates donation of Personal Protective Equipment to health care workers. We will buy one *[12 FL oz. / 355 ML hand sanitizer, 50 count disposable face masks]* from a seller in our list who charges between *[$7.50 - $10.00,$27.50 - $30.00]*.

You will receive a **$5** Amazon gift card (code to redeem it at the end of this survey).

We ensured consequentiality by verifying that `getusppe.org` had a demand for both types of PPE. Whenever a subject in our sample was randomly selected to have their donation decision implemented, we purchased the items and donated them to a hospital listed in `getusppe.org`.

In the final part of the experiment, we asked subjects questions that checked their comprehension of the experiment and their beliefs about quality differences between differently priced goods.

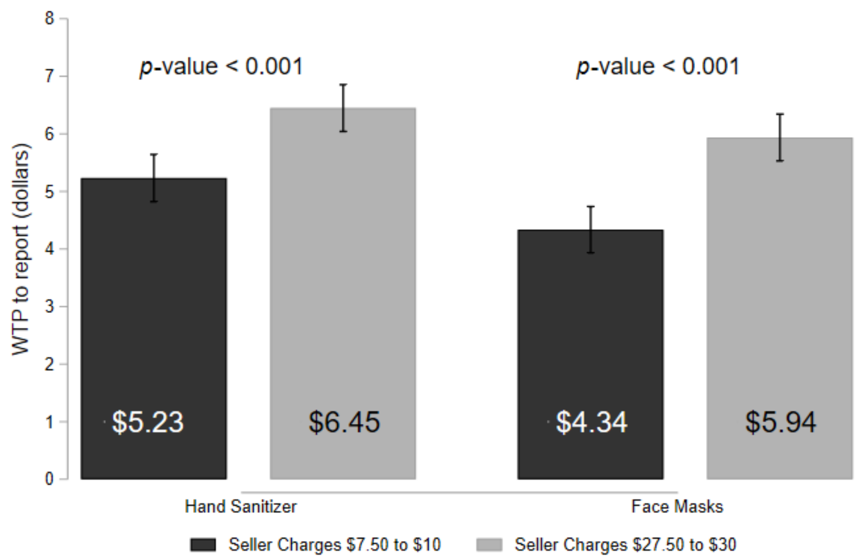
2.5 Results

2.5.1 Repugnance Toward Price Gouging

Our first goal was to test whether consumers take costly actions to oppose price gouging. Under Claim 1, one can interpret this willingness to pay to report as the level of monetary compensation under which an individual is indifferent between reporting and not reporting a seller.[30] We find that 78% of them are willing to forgo compensation to report sellers who charge in the low-price range. On average, respondents forgo over $4.8 to report sellers. Moss et al. (2020) find that most respondents report median earnings of $6-$9 per hour on Cloud Research. Our finding translates to subjects giving up 53-80% of their hourly wage.

Consistent with Claim 2, the WTPR is increasing in the ask-price. Figure 2.5 shows that increasing the price range from $7.50-10.00 to $27.50-30.00 increases the WTPR by $1.22 and $1.60 for hand sanitizer and mask, respectively. The economic significance of the treatment effect is substantial as it amounts to slightly over 20% of the pre-pandemic prices of both categories and implies an elasticity of WTPR to the ask-price of 0.17.[31]

Figure 2.5: Willingess to Pay to Report



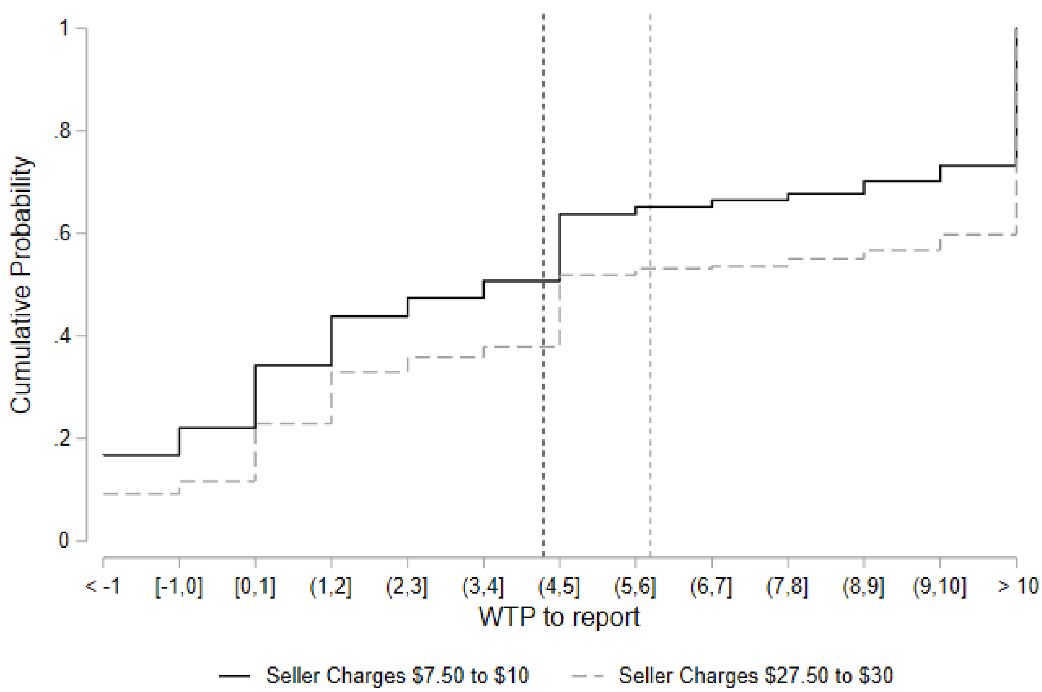
Note: Figure displays the average willingness to report sellers force price-gouging at different prices separately by PPE type with 95% confidence intervals.

30. Note that the model requires subjects do not need to view themselves as the only person who can report the seller. It only requires that people believe that by reporting, they increase the probability that a seller is reported.

31. Elasticity estimate calculated using the midpoint of the seller's price range.

The average effect underlies a more dramatic shift in the WTPR distribution. Figure 2.6 shows an increase of subjects willing to forgo the maximum potential gift card and a substantial reduction in individuals expressing indifference or a desire to pay to prevent reporting. The distributions of WTPR for both prices are statistically different (Kolmogorov-Smirnov p-value < 0.001 for face masks and for hand sanitizer). Moreover, we cannot reject that the distribution of WTPR under the high prices first and second-order stochastically dominates the distribution under the low prices (p-values of 0.8224, 0.9989 for face masks and 0.8521, 0.9986 for hand sanitizer).

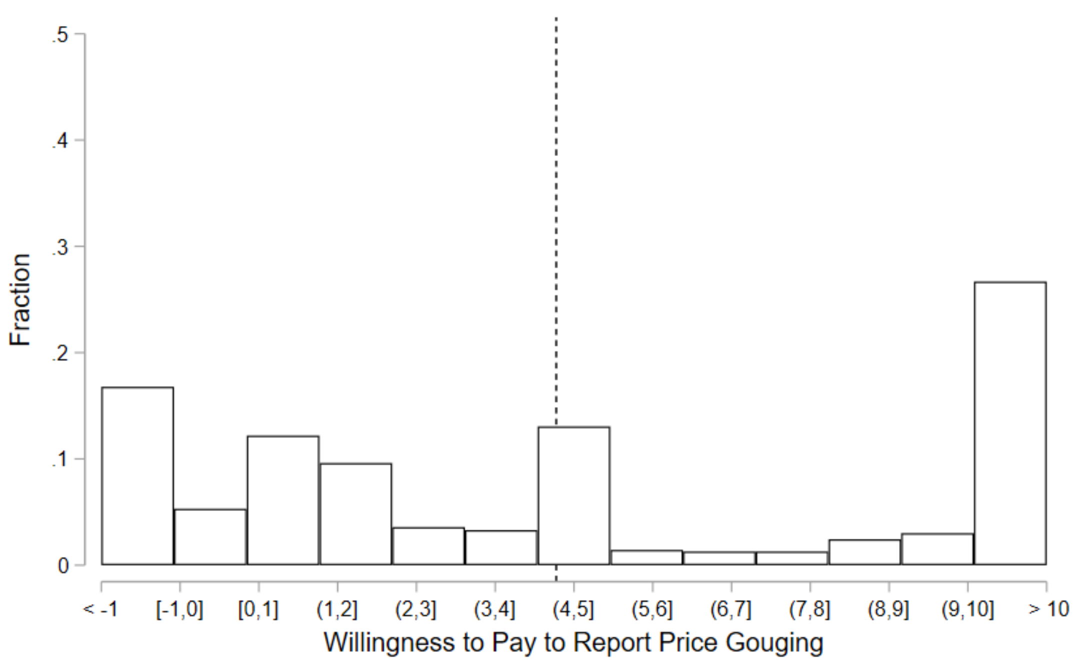
Figure 2.6: Distributions of Willingness to Pay to Report by Seller Price



Note: Figure displays presents the cumulative distribution function of willingness-to-report price gouging of either good by seller price. The vertical lines represent the average WTPR at each seller price. Kolmogorov-Smirnov p-value of 0.00003 for face masks and 0.0009 for hand sanitizer. p-values of 0.8224, 0.9989 for face masks and 0.8521, 0.9986 for hand sanitizer, for the H0 that the distribution of WTPR under high prices first and second-order stochastically dominates the distribution with low prices, using the Bootstrap tests from Abadie (2002).

Figure 2.7 also shows the distribution of WTPR to be polarized for the low-price range treatment arms; subjects have polarized preferences towards moderate price-gouging. 17% of subjects are willing to forgo one dollar or more to avoid punishing these sellers. This negative willingness to pay to report is consistent with our theoretical framework; it could be driven either by deriving negative utility from punishing sellers, a high cost of reporting, or by considering the repugnance of a given price to be much lower than the market average. We found such respondents in both price ranges, but higher-priced sellers are substantially less likely to be protected by our subjects. The polarization of the distribution of the WTPR reflects a polarization that is similar to what Elías et al. (2019) find in the context of kidney donations; some people strongly opposing the transaction and some strongly in favor of it.

Figure 2.7: Histogram of Willingness to Pay to Report at the Low Price

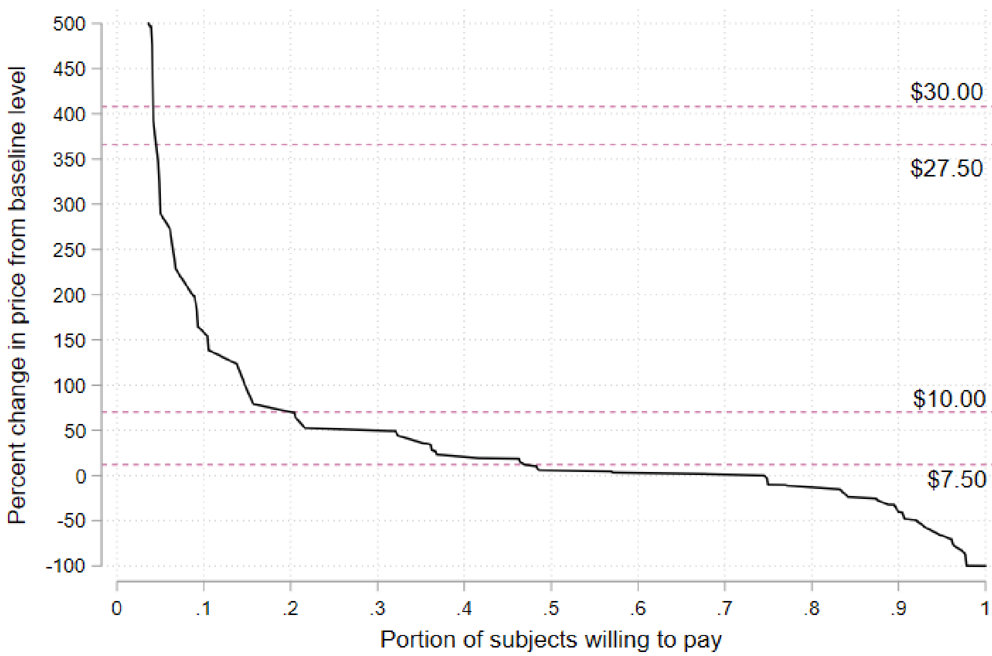


Note: This figure displays the distribution of willingness to pay to report for all subjects in the low price range treatments. The vertical line represents the average WTPR at the low-price range.

Moreover, Figure 2.8 shows the portion of subjects willing to pay for either type of

PPE at different percent changes from the December price collected in the second step of the experiment. Almost half of the subjects are willing to buy the goods from "low-price" sellers, while at least five percent are still willing to buy from "high-price" sellers.[32] Since 50% of subjects are willing to purchase PPE at prices in the lower price range, the decision to punish sellers implies that subjects find these transactions repugnant (Roth, 2007).[33] That is, subjects prevent voluntary transactions between third-parties.

Figure 2.8: Willingness to Pay for Personal Protective Equipment



Note: This figure shows the portion willing to pay for either type of PPE at% changes from pre-crisis prices. Horizontal lines denote the treatment price ranges.

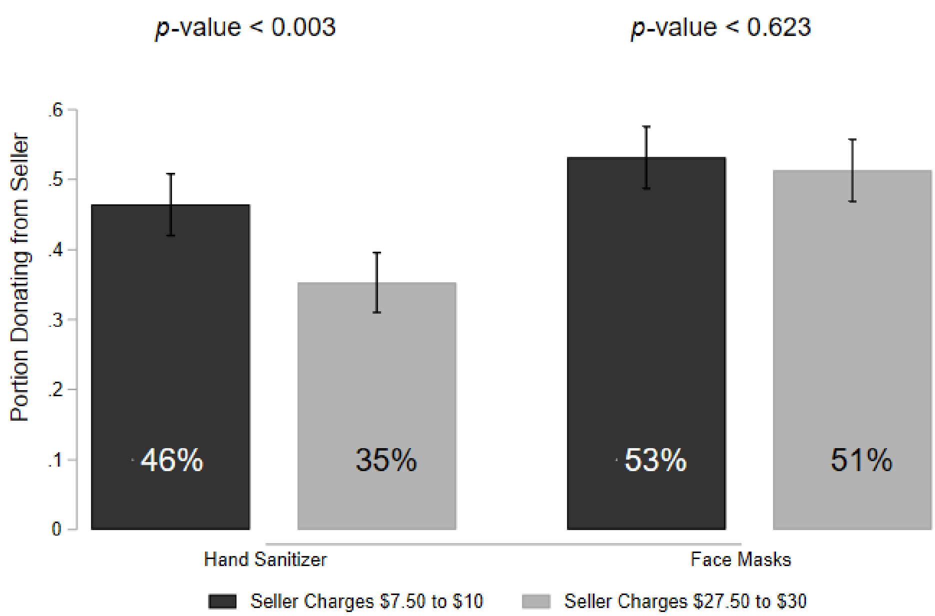
32. With the caveat, as we argued above, that our measure of WTP for the PPE is biased downwards, since we elicit it with a procedure similar to a first-price auction.

33. As we argued above, it is unlikely that individuals receive any direct benefit (other than moral benefit) from reporting sellers, since we match them with a random seller chosen from a large pool. This means that they cannot claim any refund or expect to face lower prices or search costs in the future because of this decision.

2.5.2 Underlying Motives

Under Claim 3 of our model, individuals that derive external benefits from PPE consumption derive positive surplus from donations of the product. Consistent with this, a significant portion of subjects are willing to forgo the five dollars to have us donate the PPE in every treatment. Using variation in the seller's price in the donation experiment, we find that the mechanism driving the repugnance towards price-gouging is good specific. The donation rates for subjects considering hand-sanitizer decrease by 30% when we purchase the good from a higher priced seller. Conversely, the subjects consider face masks are uninfluenced by seller price (see Figure 2.9). In other words, we find evidence of distaste for firm profits with hand sanitizer, but not face masks. This result is striking, since the willingness to pay to report face masks was at least as responsive to seller price as hand sanitizer.

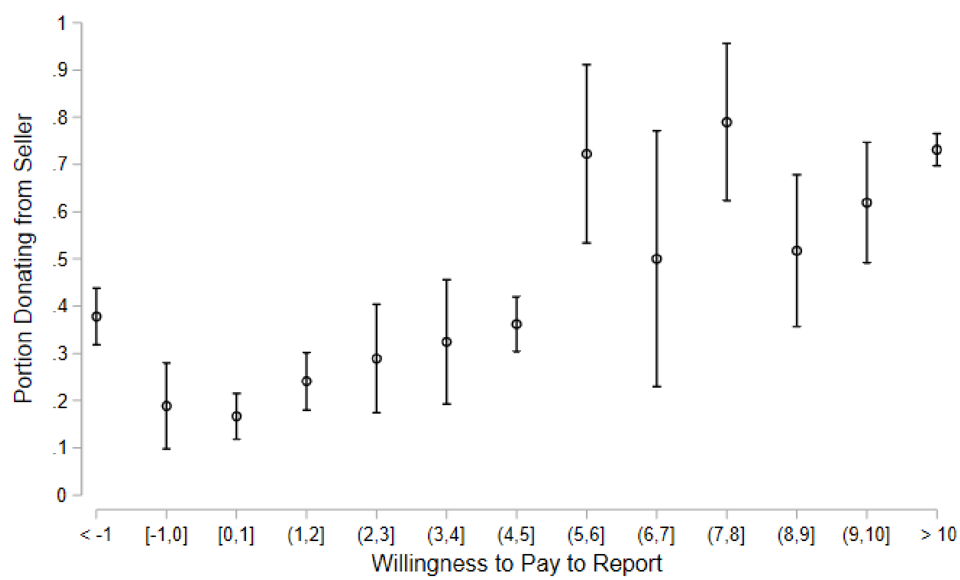
Figure 2.9: Propensity to Donate PPE from Price-Gougers



Note: This Figure displays the average willingness to report sellers force price-gouging at different prices separately by PPE type with 95% confidence intervals.

There are many potential explanations for the difference in mechanisms between products. This could be driven by a difference in how individuals perceive masks to be of a higher necessity than hand sanitizer (see the discussion in Tobin (1970) of how basic necessities might trigger distributional concerns). It could also be the case that masks are perceived to have a higher spillover on other individuals—and so the "external" component of repugnance is higher. Since all subjects completed both tasks and we estimated the WTPR within-subject, we can use the within-person relationship between these choices to check for consistency between donation and reporting decisions.

Figure 2.10: Relationship between Willingness to Report and Propensity to Donate



Note: This Figure plots the average portion of subjects choosing to donate PPE within every willingness to report bin. This figure pools both seller prices and types of PPE and reports 95% confidence intervals.

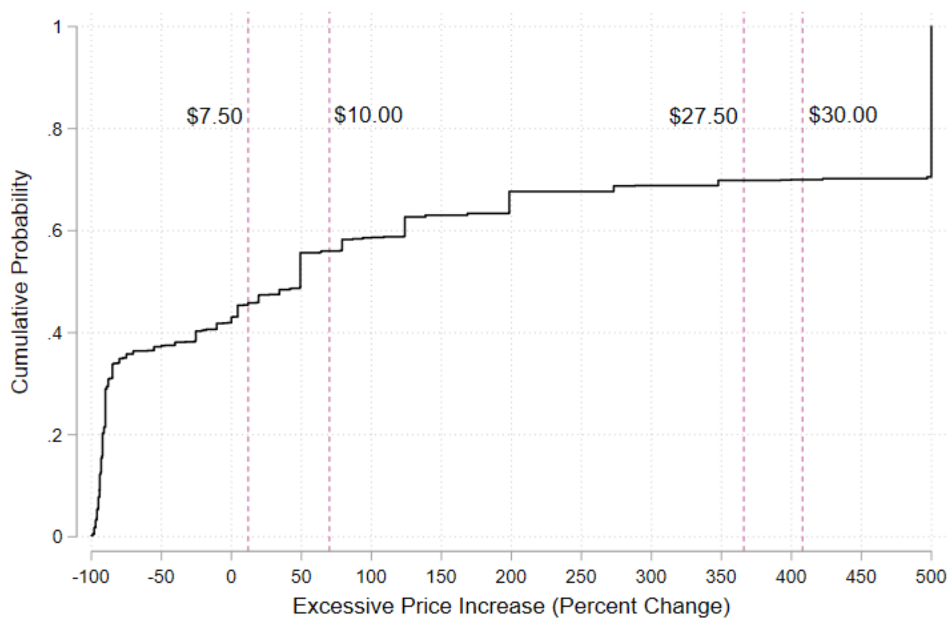
Nearly 50% of subjects who were willing to pay positive amounts to report sellers were also willing to donate. Figure 2.10 reports a generally positive association between WTPR and donations. This positive correlation is expected from our model, since individuals who derive a high benefit from others' consumption should both be more willing to donate and have a higher repugnance; which would manifest as a higher willingness to pay to report.

A notable exception is that subjects who are willing to pay to avoid reporting sellers have donation rates twice as large as those who express a WTPR of zero ($p < 0.001$). Their donation rate is less than the average of all subjects who are willing to pay to report price-gouging, and comparable to subjects willing to pay \$2 to \$5 to report sellers.

2.5.3 Heterogeneity

The heterogeneity in WTPR across individuals' perception of "excessiveness" is reported in Figure 2.11. Figure 2.11 displays the CDF of self-reported excessive prices for either type of PPE at different % changes from pre-crisis prices. Only 40% of respondents would consider prices in the lower price range excessive while more than 70% deem prices in the higher price range excessive.

Figure 2.11: Excessive Price of Personal Protective Equipment



Note: This figure displays the CDF of self-reported excessive prices for either type of PPE at percent changes from pre-crisis prices. Vertical lines denote the potential seller price ranges. Data are winsorized at the 99th percentile as mentioned in our pre-analysis plan.

We find that a non-negligible portion of the sample finds pre-pandemic prices excessive—close to 40%. One possible driver of this finding is the social norm or preference of individuals for having firms engage in pro-social activities during emergencies (which may include lowering prices). Indeed, Marcelo et al. (2020) document many different examples of firms participating in non-profit or Corporate Social Responsibility (CSR) initiatives during crises.[34] This finding could also be due to the hypothetical and subjective nature of the question of what "excessive" means. Figure 2.12 shows that the WTPR is higher when the price range is considered to be excessive ex-ante. The WTPR might provide a revealed-preference version of "excessive" and be more suitable for policy (e.g., to define which price increases are considered to be price-gouging). Indeed, there is a close relationship between both measures, since individuals who consider a price range excessive have a 32% higher WTPR than those who don't.

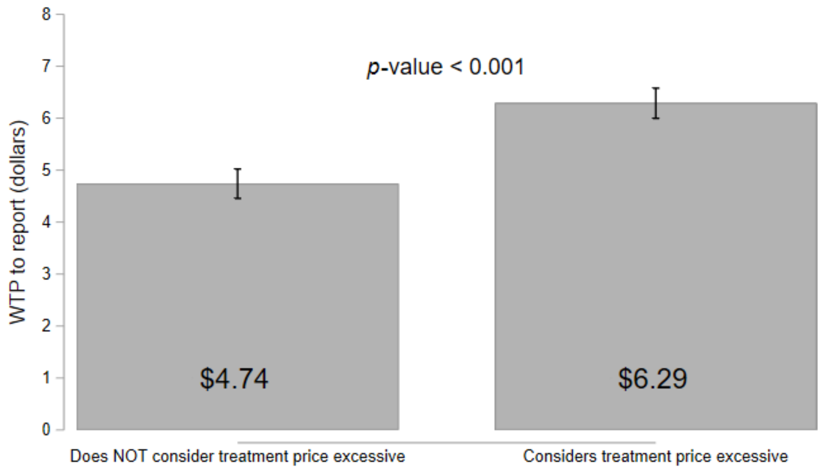
2.6 Generalizability and Robustness

We use List (2020)'s SANS conditions to understand the experiment's generalizability to the entire United States. We selected our subjects to match the U.S. on race, Hispanic origin, age, and gender. However, the survey over samples subjects with a high-school education and under samples subjects with less than high school or more than a four-year degree. We reweight our data to match U.S. population moments to learn about the external validity of our estimates (Hotz et al., 2005). Table 2.5 shows that reweighting does not materially change the results.

The compliance rate after the randomization is 98%. There are also no motivational or incentive differences across treatments that materially affect attrition. Nevertheless, we use the non-parametric approach in Manski (1989) to derive treatment effect bounds with our

34. He and Harris (2020) give examples of CSR in the context of Covid. Similarly, Uber offered free rides to passengers during natural disasters in the example above rather than offering rides at non-surge prices.

Figure 2.12: WTPR by whether Seller Price is Considered Excessive



Note: This figure displays the average WTPR split by whether the subject reported that they found values in the seller's price range excessive and 95% confidence intervals. Estimates pool subjects across all treatments and exclude subjects who did not report a WTP or excessive price for the PPE considered in their treatment.

data. Our results persist, with less precision, when using the bounding approach.

Regarding the naturalness of the experiment, we use a framed field experiment (Harrison and List, 2004). Price-gouging legislation activates during declared states of emergency. While atypical, we are operating in precisely the setting to which we wish to generalize. The text analysis of our sample of actual price gouging complaints (Section 2.2) shows that complaints about face masks and hand sanitizers were common. The iMPL may be unnatural to subjects, but we are comparing choices made in the iMPL to consequential choices made by thousands of individuals outside of the experiment.

Moreover, Berry et al. (2020) shows that choices made using within-person elicitations are congruent with decisions in more natural take-it-or-leave-it offers. Since the experiment takes place online at the subject's own pace, subjects are free to seek information that would aid in their decision-making. The donation experiment mimics actions taken by private companies during other natural disasters (Uber, 2016). We view our WTPR and mechanism

Table 2.5: Main Results for Calibrated Sample to Match U.S. Adults

	(1) WTR	(2) WTR	(3) WTR	(4) WTR	(5) Donation	(6) Donation	(7) Donation	(8) Donation
Seller Charges \$27.50 to \$30	1.246**	1.142*			-0.0776	-0.0448		
	(0.461)	(0.449)			(0.0480)	(0.0487)		
Would buy at treatment price			-1.072*					
			(0.455)					
Considers treatment price excessive				1.646***				
				(0.321)				
$\mathbb{1}\{WTR > 0\}$							0.324***	
							(0.0617)	
$\mathbb{1}\{WTR = 0\}$							0.191*	
							(0.0752)	
$\mathbb{1}\{WTR = -1\}$								0.191*
								(0.0752)
$\mathbb{1}\{WTR = 1\}$								0.00590
								(0.0712)
$\mathbb{1}\{WTR = 2\}$								0.0491
								(0.0704)
$\mathbb{1}\{WTR = 3\}$								0.172
								(0.103)
$\mathbb{1}\{WTR = 4\}$								0.226*
								(0.112)
$\mathbb{1}\{WTR = 5\}$								0.173*
								(0.0730)
$\mathbb{1}\{WTR = 6\}$								0.616***
								(0.122)
$\mathbb{1}\{WTR = 7\}$								0.352
								(0.182)
$\mathbb{1}\{WTR = 8\}$								0.663***
								(0.103)
$\mathbb{1}\{WTR = 9\}$								0.353**
								(0.121)
$\mathbb{1}\{WTR = 10\}$								0.477***
								(0.110)
$\mathbb{1}\{WTR = 11\}$								0.557***
								(0.0650)
Constant	5.297***	4.443***	5.833***	4.688***	0.442***	0.540***	0.166**	0.166**
	(0.333)	(0.320)	(0.248)	(0.226)	(0.0343)	(0.0349)	(0.0588)	(0.0588)
N	695	696	732	1391	692	694	1386	1386

Note: This table replicates the main results from the paper after re-weighting observations to match the marginal distribution of gender, age, ethnic affinity, education and income. Robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

insights as wave 1 insights in the langauge of List (2020). Further work should attempt to understand the WTPR for goods that do not have positive externalities and focus on understanding what drives the differences in mechanisms across goods.

Regarding internal validity, there are four potential confounders to our results. First, there might be experimenter demand effects that incentivize individuals to align their responses to what they perceive to be our desired results. To reduce this possibility, we provide full anonymity to our participants (de Quidt et al., 2019). We coded the survey to embed bonus payments to avoid asking for any identifying information.[35] Moreover, the heterogeneity observed Figure 2.9 suggests that any experimenter demand effect would need to be good specific, which is unlikely. Second, the treatment might be too subtle for individuals to notice. We asked individuals an attention question at the end of the survey, in which they had to report the price range that they were assigned.

Table 2.6: Treatment Effect on Attention

	(1) Attention	(2) Attention	(3) Attention	(4) Attention
Seller Charges $27.50 to $30	-0.196***	-0.196***	-0.190***	-0.214***
	(0.0187)	(0.0187)	(0.0188)	(0.0265)
Face Masks		0.000645	-0.00182	0.0221
		(0.0186)	(0.0185)	(0.0179)
Seller Charges $27.50 to $30 × Face Masks				0.0482
				(0.0374)
Constant	0.944***	0.944***	0.797***	0.786***
	(0.00871)	(0.0134)	(0.0629)	(0.0630)
Controls	NO	NO	YES	YES
R2	0.074	0.074	0.109	0.110
Observations	1391	1391	1391	1391

Note: Table displays the effect of treatments on the propensity correctly answer the attention question. Omitted category is hand sanitizer sold for $7.50 to $10.00. Controls include race, sex, income, and education. State laws is an indicator equal to 1 if the subject's state has laws against price gouging. Heteroskedasticity robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 2.6 shows that individuals in the high price range tend to misremember the price range that they were given; that is, they report incorrectly that they were assigned the lower price range. This means that our results, if anything, are biased downwards, since some people in the upper price range believe that they were assigned the lower prices. Indeed, Table 2.7 displays the treatment effects for the willingness to pay to report and donations for the overall sample and only attentive subjects. When we restrict the analysis to attentive

35. Many field experiments compensate participants by sending gift cards to their email address.

Table 2.7: Treatment Effects on Attentive Subjects

	(1) WTP	(2) WTP	(3) Donate	(4) Donate
Seller Charges \$27.50 to \$30	1.621***	2.093***	-0.0173	0.0111
	(0.346)	(0.380)	(0.0380)	(0.0416)
Face Masks	-0.916***	-0.929**	0.0661*	0.0624
	(0.348)	(0.360)	(0.0379)	(0.0390)
Seller Charges \$27.50 to \$30 × Face Masks	-0.405	-0.359	-0.0941*	-0.117**
	(0.492)	(0.538)	(0.0531)	(0.0582)
Constant	5.232***	5.160***	0.464***	0.468***
	(0.248)	(0.260)	(0.0267)	(0.0277)
Elastcity Estimate	1.01	1.3	-.01	.01
R2	0.030	0.048	0.019	0.019
Observations	1391	1177	1386	1172

Note: Table displays the effect of treatments on the WTP to report and propensity to donate. Omitted category is hand sanitizer sold for \$7.50 to \$10.00. Odd Columns include the full sample of subjects. Even columns drop subjects who answered the attention question incorrectly. Controls include race, sex, income, and education. State laws is an indicator equal to 1 if the subject's state has laws against price gouging. Heteroskedasticity robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

subjects, we find a higher WTPR and a more substantial reduction in donations when considering high-priced face masks.

Third, individuals might perceive that products with higher prices differ in other ways as well from products with lower prices (e.g., differences in quality, shipping dates, etc.). We tell individuals that our algorithm has found products in the previous weeks with prices from \$5 to \$50 with similar shipping dates. At the end of the survey, we ask the subjects whether they agree with the statement that products in the upper price range have a higher quality than products in the lower price range. Table 2.8 shows that treatment status has mostly insignificant impact on quality beliefs.

Table 2.8: Treatment Effect on Higher Quality Belief

	(1) Higher Quality	(2) Higher Quality	(3) Higher Quality	(4) Higher Quality
Seller Charges \$27.50 to \$30	-0.0336	-0.0336	-0.0379*	-0.0161
	(0.0220)	(0.0220)	(0.0211)	(0.0305)
Face Masks		0.0228	0.0239	0.00234
		(0.0220)	(0.0211)	(0.0305)
Seller Charges \$27.50 to \$30 × Face Masks				-0.0433
				(0.0417)
Constant	0.231***	0.220***	0.317***	0.327***
	(0.0160)	(0.0194)	(0.0656)	(0.0658)
Controls	NO	NO	YES	YES
R2	0.002	0.002	0.123	0.123
Observations	1391	1391	1391	1391

Note: Table displays the effect of treatments on the propensity to claim that higher priced PPE is higher quality. Omitted category is hand sanitizer sold for \$7.50 to \$10.00. Controls include race, sex, income, and education. State laws is an indicator equal to 1 if the subject's state has laws against price gouging. Heteroskedasticity robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Lastly, individuals might also be repugnant to accepting money in exchange for reporting a seller. For instance, Roth (2007) argues that some exchanges become repugnant when money is incorporated into the transaction. While we cannot rule this out, there is at least a partial rate of substitution between cash payments and reporting or donating since WTPR and donation rates are responsive to our treatment. Moreover, this would only bias our estimates downwards since higher cash payments would also entail a higher 'cash repugnance'. An individuals valuation from reporting sellers would thus be higher than what they reveal through cash incentives.

2.7 Conclusions

In this paper we propose an incentivized reporting experiment (IRE) to quantify the repugnance to price-gouging and unpack its mechanisms. Using our theoretical model, we argue that reporting a seller for price-gouging contains information about repugnance to this transaction, as well as expected benefits from punishing the seller. Based on this, the IRE elicits the willingness to pay of individuals to report a seller to the authorities for price-gouging. While there is some correlation between the elicited WTPR and stated-preference measures of whether prices are considered to be excessive, our revealed-preference approach might be more useful to determine the price ranges sanctioned by price-gouging laws. Beyond this application, IREs could be used to study repugnance towards activities in which enforcement requires that illicit activity is reported to the authorities.

We show that most individuals value reporting price increases of face masks and hand sanitizers during the first wave of COVID-19, although there is some polarization. Individuals also respond to the seller's price and increase their willingness to pay to report when facing more expensive sellers. The documented measure implies opposition to transactions that some participants would find beneficial and thus presents a consequential example of repugnant transactions in the field. Our results are consistent with prior studies on repugnance, such as Elías et al. (2019), who also find that individuals are willing to tolerate inefficiencies in order to reduce repugnance and that they have polarized preferences. Moreover, the experiment shows that raising the price of essential products during emergencies has economically significant negative externalities on third-parties.

A choice between a $5 gift card and having us donate an item of PPE purchased from a price-gouger clarifies the underlying motivation behind the opposition to large price increases during emergencies. We find evidence for distaste for seller profits in the case of hand sanitizers but a higher priority for others' consumption when it comes to face masks. The fact that individuals may obtain negative payoffs from profits in the case of some products suggests an additional welfare cost of policies such as subsidies—that potentially increase profits—versus price controls. Moreover, further research should understand what drives the heterogeneity across products.

2.8 Appendix

2.8.1 Theoretical Framework

The expected value of both accepting and reporting an offer z is:

$$v^{r,a}(z) = \beta\kappa - c_r + \widetilde{u}(z) + \frac{1}{M}\underbrace{\left[-\underbrace{\sigma^a(z)R(z)}_{\text{The other consumer accepts}} - \underbrace{(1-\sigma^a(z))\mathbb{E}R}_{\text{The other consumer searches}}\right]}_{\text{Consumers matched to the same seller}}$$
$$+\frac{M-1}{M}\int_{\underline{p}}^{\overline{p}}\underbrace{\left[-\underbrace{\sigma^a(p)R(p)}_{\text{Accepts}} - \underbrace{(1-\sigma^a(p))\mathbb{E}R}_{\text{Searches}} + \underbrace{\sigma^r(p)\beta\kappa}_{\text{Reports}}\right]}_{\text{Consumers matched to different sellers}} dF(p)$$

The value of reporting the offer and not accepting it (i.e., searching for other offers) is:

$$v^{r,s}(z) = \beta\kappa - c_r - c_s + \frac{1}{M}\underbrace{\left[\mathbb{E}\widetilde{u} - \underbrace{\sigma^a(z)R(z)}_{\text{The other consumer accepts}} - \underbrace{(1-\sigma^a(z))\mathbb{E}R}_{\text{The other consumer searches}}\right]}_{\text{Consumers matched to the same seller}}$$
$$+\frac{M-1}{M}\int_{\underline{p}}^{\overline{p}}\underbrace{\left[-\underbrace{\sigma^a(p)R(p)}_{\text{Accepts}} - \underbrace{(1-\sigma^a(p))\mathbb{E}R}_{\text{Searches}}\right.}_{\text{Consumers matched to different sellers}}$$
$$\underbrace{\left.+\underbrace{\sigma^r(p)(\beta\kappa+\mathbb{E}\widetilde{u})}_{\text{Reports}} + \underbrace{(1-\sigma^r(p))\left(\frac{M-2}{M-1}\mathbb{E}\widetilde{u} + \frac{1}{M-1}\widetilde{u}(p)\right)}_{\text{Does not report}}\right]}_{\text{Consumers matched to different sellers}} dF(p)$$

The value of accepting an offer and not reporting it is:

$$v^{n,a}(z) = \widetilde{u}(z) + \frac{1}{M}\underbrace{\left[-\underbrace{\sigma^a(z)R(z)}_{\text{The other consumer accepts}} - \underbrace{(1-\sigma^a(z))\mathbb{E}R}_{\text{Searches}} + \underbrace{\sigma^r(z)\beta\kappa}_{\text{Reports}}\right]}_{\text{Consumers matched to the same seller}} +$$

$$\underbrace{\frac{M-1}{M}\int_{\underline{p}}^{\overline{p}}\left[-\underbrace{\sigma^a(p)R(p)}_{\text{Accepts}} - \underbrace{(1-\sigma^a(p))\left(\frac{M-2}{M-1}\mathbb{E}R + \underbrace{\frac{1}{M-1}R(z)}_{\text{Might match } z}\right)}_{\text{Searches}} + \underbrace{\sigma^r(p)\beta\kappa}_{\text{Reports}}\right]}_{\text{Consumers matched to different sellers}} dF(p)$$

Finally, the value of neither accepting nor reporting an offer is:

$$v^{n,s}(z) = -c_s + \frac{1}{M}\underbrace{\left[\mathbb{E}\widetilde{u} - \underbrace{\sigma^a(z)R(z)}_{\text{The other consumer accepts}} - \underbrace{(1-\sigma^a(z))\mathbb{E}R}_{\text{Searches}} + \underbrace{\sigma^r(z)\beta\kappa}_{\text{Reports}}\right]}_{\text{Consumers matched to the same seller}}$$

$$+\frac{M-1}{M}\int_{\underline{p}}^{\overline{p}}\underbrace{\left[-\underbrace{\sigma^a(p)R(p)}_{\text{Accepts}} - \underbrace{(1-\sigma^a(p))\left(\frac{M-2}{M-1}\mathbb{E}R + \underbrace{\frac{1}{M-1}R(z)}_{\text{Might match } z}\right)}_{\text{Searches}}\right.}_{\text{Consumers matched to different sellers}}$$

$$\underbrace{\left.+\underbrace{\sigma^r(p)(\beta\kappa + \mathbb{E}\widetilde{u})}_{\text{Reports}} + \underbrace{(1-\sigma^r(p))\left(\frac{M-2}{M-1}\mathbb{E}\widetilde{u} + \frac{1}{M-1}\widetilde{u}(p)\right)}_{\text{Does not report}}\right]}_{\text{Consumers matched to different sellers}} dF(p)$$

Then, the value of reporting versus not reporting satisfies:

$$\begin{aligned} v^r(z) - v^n(z) &= v^{r,a}(z) - v^{n,a}(z) = v^{r,s}(z) - v^{n,s}(z) \\ &= \beta\kappa\left(1 - \frac{\sigma^r(z)}{M}\right) - c_r + \left(\frac{R(z) - \mathbb{E}R}{M}\right)(1 - \mathbb{E}\sigma^a) \end{aligned} \tag{2.5}$$

And the value of accepting an offer versus searching:

$$\begin{aligned} v^a(z) - v^s(z) &= v^{r,a}(z) - v^{r,s}(z) = v^{n,a}(z) - v^{n,s}(z) \\ &= c_s + \widetilde{u}(z) - \mathbb{E}\widetilde{u} + \frac{\mathrm{Cov}(\sigma^r, \widetilde{u})}{M} \end{aligned} \tag{2.6}$$

An equilibrium is given by functions σ^a and σ^r with range in $[0,1]$ such that $\sigma^r(z) = 0$ implies $v^r(z) - v^n(z) \le 0$, $\sigma^r(z) = 1$ implies $v^r(z) - v^n(z) \ge 0$ and $\sigma^r(z) \in (0,1)$ implies $v^r(z) - v^n(z) = 0$. σ^a has to be consistent in a similar way with $v^a(z) - v^s(z)$ and they both have to be consistent with $\mathrm{Cov}(\sigma^r, \widetilde{u})$ and $\mathbb{E}\sigma^a$. In other words, we proceed by first finding conditional $\sigma^r(z; \mathbb{E}\sigma^a)$ and $\sigma^a(z; \mathrm{Cov}(\sigma^r, \widetilde{u}))$ and then showing that there exist equilibrium values A^* and B^* such that
$\mathrm{Cov}(\sigma^r(z; A^*), \widetilde{u}(z)) = B^*$ and $\mathbb{E}[\sigma^a(z; B^*)] = A^*$.

We first find the conditional equilibrium probability of reporting, $\sigma^r(z; \mathbb{E}\sigma^a)$. Fix $\mathbb{E}\sigma^a$. Define $f(z,s) = \beta\kappa(1 - s/M) - c_r + \left(\frac{R(z) - \mathbb{E}R}{M}\right)(1 - \mathbb{E}\sigma^a)$. Note that $f(z,s)$ is increasing in z (strictly if $\mathbb{E}\sigma^a < 1$) and strictly decreasing and continuous in s. The functions $f(z,0)$ and $f(z,1)$ bound $v^r(z) - v^n(z)$ since $v^r(z) - v^n(z) = f(z, \sigma^r(z)) \in [f(z,1), f(z,0)]$. For any z such that $f(z,0) \le 0$ (if any such z exists), any equilibrium $\sigma^r(z)$ has to be zero, since otherwise $\sigma^r(z) = s^* > 0$ would imply $v^r(z) - v^n(z) = f(z, s^*) < f(z,0) \le 0$, which contradicts the equilibrium conditions above. For any z such that $f(z,1) < 0 < f(z,0)$, any equilibrium has to satisfy $\sigma^r(z) \in (0,1)$, so $\sigma^r(z) = s^*$ for an s^* such that $f(z, s^*) = 0$. A unique s^* exists since f is strictly decreasing and continuous in s. Lastly, for any z such that $f(z,1) \ge 0$, any equilibrium $\sigma^r(z)$ has to be one, since $\sigma^r(z) = s^* < 1$ would imply $v^r(z) - v^n(z) = f(z, s^*) > f(z, 1) \geq 0$. Hence, the unique equilibrium (conditional) probability of reporting is $\sigma^r(z) = \max\{\min\{\sigma^*(z), 1\}, 0\}$, where:

$$\sigma^*(z) = \frac{M}{\beta\kappa}\left[\beta\kappa - c_r + \left(\frac{R(z) - \mathbb{E}R}{M}\right)(1 - \mathbb{E}\sigma^a)\right] \tag{2.7}$$

Now we find the equilibrium (conditional) probability of accepting, $\sigma^a(z; \mathrm{Cov}(\sigma^r, \widetilde{u}))$. If there exists a reservation price $r \in [\underline{p}, \overline{p}]$ such that $v^a(r) - v^s(r) = 0$, then it is unique, since $\widetilde{u}(z)$ is strictly decreasing. Define the reservation price r to be equal to this unique root of $v^a(r) - v^s(r)$, if it exists, and let $r = \overline{p}$ if $v^a(\overline{p}) - v^s(\overline{p}) > 0$ and $r = \underline{p}$ if $v^a(\underline{p}) - v^s(\underline{p}) < 0$. Note that $v^a(z) - v^s(z) >=< 0 \iff z <=> r$, so $\sigma^a(z; \mathrm{Cov}(\sigma^r, \widetilde{u}))$ is 1 for $z \leq r$ and 0 for $z \geq r$.

Note that $\mathbb{E}[\sigma^a(z)] = \Pr(z \leq r) = F(r)$, so consistency of $\mathbb{E}\sigma^a$ and $\mathrm{Cov}(\sigma^r, \widetilde{u})$ reduces to finding consistency of $F(r)$ and $\mathrm{Cov}(\sigma^r, \widetilde{u})$. Equations (2.6) and (2.7) pin down the equilibrium. The reservation price r is a weakly increasing function of $\mathrm{Cov}(\sigma^r, \widetilde{u})$; more negative values of this covariance (which is negative since $\widetilde{u}$ is decreasing and σ^r is increasing) make the right-hand side of Equation (2.6) lower. Hence, $F(r(\mathrm{Cov}(\sigma^r, \widetilde{u})))$ is also an increasing function of $\mathrm{Cov}(\sigma^r, \widetilde{u})$. Call this function $\gamma(\mathrm{Cov}(\sigma^r, \widetilde{u}))$, which is between 0 and 1. If $\mathrm{Cov}(\sigma^r, \widetilde{u}) \leq \mathbb{E}\widetilde{u} - \widetilde{u}(\underline{p}) - c_s \leq 0$ then $\gamma = 0$. If $\mathrm{Cov}(\sigma^r, \widetilde{u}) \geq \mathbb{E}\widetilde{u} - \widetilde{u}(\overline{p}) - c_s$ (if this number is negative) then $\gamma = 1$. When $\mathrm{Cov}(\sigma^r, \widetilde{u})$ is in between these bounds, γ is strictly increasing.

Plugging in $\mathbb{E}\sigma^a = F(r)$ in Equation (2.7) we can see that higher $F(r)$ reduces $\sigma^*(z)$ for z bigger than z^e, where $R(z^e) = \mathbb{E}R$, and increases $\sigma^*(z)$ for $z < z^e$, so $\sigma^r(z)$ becomes flatter. Hence, higher $F(r)$ increases the covariance $\mathrm{Cov}(\sigma^r, \widetilde{u})$ (makes it less negative).[36] Call this function $\delta(F(r))$, which takes non-positive values. Note that $\delta(1) = 0$ and $\delta(F(r)) \leq 0$ for any $F(r) < 1$. An equilibrium is thus a value F^* such that both curves intersect;

36. To see why, let $F(r') > F(r)$. We just argued that $\sigma^r(z; F(r)) - \sigma^r(z; F(r'))$ is an increasing function of z. To show that $\mathrm{Cov}(\sigma^r(z; F(r)), \widetilde{u}(z)) < \mathrm{Cov}(\sigma^r(z; F(r')), \widetilde{u}(z))$, note that $\mathrm{Cov}(\sigma^r(z; F(r)), \widetilde{u}(z)) - \mathrm{Cov}(\sigma^r(z; F(r')), \widetilde{u}(z)) = \mathrm{Cov}(\sigma^r(z; F(r)) - \sigma^r(z; F(r')), \widetilde{u}(z)) < 0$. This is negative since it's the covariance between an increasing and a decreasing function of z.

$\gamma(\delta(F^*)) = F^*$. We have two cases. If $\delta(0) < \mathbb{E}\widetilde{u} - \widetilde{u}(\underline{p}) - c_s$, then $F^* = 0$ is an equilibrium, since $\gamma(\delta(0)) = 0$. There could be additional equilibria if $\gamma(0) = 1$. In the second case, if $\delta(0) > \mathbb{E}\widetilde{u} - \widetilde{u}(\underline{p}) - c_s$, then there exists a single equilibrium since $\delta \in [0, 1]$ and γ has to cross δ at some point to the right of $\delta(0)$.

2.8.2 Product Tracking Algorithm

To track goods and prices for our survey respondents we used the Rainforest API. It allowed us to get real-time data on availability, prices and comments on all products that are listed in the queries to "hand sanitizer" and "face mask".

The steps of the algorithm were:

1. Get the list of products that appear in the search results for the Hand Sanitizer and Face mask categories.[37]

2. Get information for each product: price, image, description, shipping date, etc.

3. Run an image classification algorithm to select which products were actually hand sanitizers and face masks

4. Process the text in the title, product description and product dimensions with regular expressions to extract and parse the number of units (fl oz, count, etc.)

We collected search results on 7 dates, covering the 2 days that our survey lasted and 2 weeks before and after our experiment. We collect prices, listing titles and product images for all searches. The output from these queries included some "false positive" results, that is, not everything was truly one of the products we cared about. Since many products are

37. Hand sanitizers can be found in product category 2265897011; see `https://www.amazon.com/Hand-Sanitizers/b?ie=UTF8&node=2265897011`. Likewise, face masks correspond to product catecories 6125377011, 8404646011 and 17864516011.

Table 2.9: Extracted Title Features

Face Mask	'cloth', 'Surgical', 'Dust', 'respirator', 'dust', 'reusable'
Hand Sanitizer	'hand', 'gel', 'Purell', 'WIPES', 'TISSUES', 'paper', 'glo', 'GERM', 'lamp', 'uv', 'ULTRAVIOLET', 'IODINE', 'cotton', 'lotion', 'spray', 'air', 'holder', 'dispenser', 'soap'

advertised in multiple search categories (e.g., soaps in the hand-sanitizer section), to avoid tracking and reporting incorrect items we classified 1200 results for"face mask" and 500 results for "hand sanitizer" with the help of Amazon MTurk workers to identify surgical face masks and alcohol based hand sanitizer gel. We used 3 labels to classify face masks: surgical masks, N-95 and not a mask. We used a binary label for hand sanitizer. These examples were then used to train a neural network classifier on PyTorch that used product images and text features from the product title as input to identify items of interest.

We used the pre-trained resnet50 model available in Torchvision to extract features from product images (see He et al. (2016)). To this convolutional model, we added two extra linear layers that allowed us to incorporate a vector of zeros and ones that identified the presence of particular words in the product title. The word-features used for each product model can be found in Table 2.9. During the learning step, only the last linear layer of the resnet50 model and the two extra layers had their weights updated to fully take advantage of knowledge already incorporated in the pre-trained model. The trained model had an out-of-sample accuracy of 0.95 and cross-entropy loss of 0.23 for Hand Sanitizers while the respective quantities were 0.97 and 0.0957 for Masks.

Afterwards, we collected more detailed product characteristics from the filtered results, such as shipping dates, stock availability, product description and dimensions. As detailed on step 4 above, we used this information to convert prices into common units.

2.8.3 Survey

Demographic Questions

1. What is your U.S. ZIP code?

2. What is your year of birth?

3. What is the highest level of school you have completed or the highest degree you have received?

    - Less than high school degree
    - High school graduate (high school diploma or equivalent including GED)
    - Some college but no degree
    - Associate degree in college (2 year)
    - Bachelor's degree in college (4 year)
    - Master's degree
    - Doctoral degree
    - Professional degree (JD, MD)

4. Choose one or more races/ethnicities that you consider yourself to be:

    - White or European American
    - Black or African American
    - Hispanic or Latino
    - Asian or Asian American
    - Other:

5. What is your approximate household annual income? Please indicate the answer that includes your entire household income in 2019 before taxes

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

6. What is your sex? Male/Female

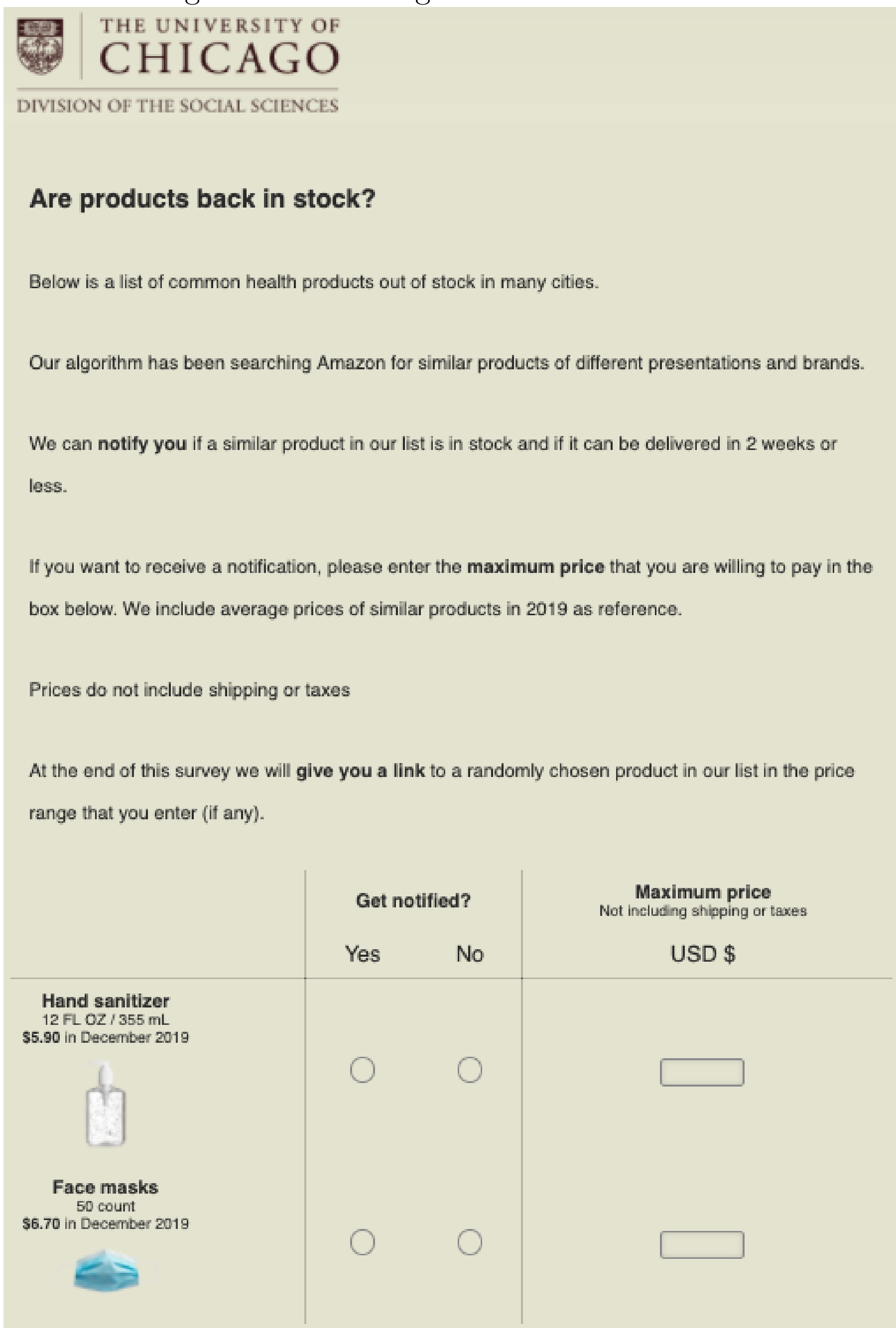
7. Have you purchased anything on Amazon in the last month? Yes/No

8. Do you have Amazon Prime? Yes/No

9. Have you bought online or in stores any of the following in 2020? Please select all that apply:

    - Hand sanitizer
    - Face masks
    - None of the above

Figure 2.13: Willingness to Track the Items



THE UNIVERSITY OF CHICAGO

DIVISION OF THE SOCIAL SCIENCES

Are products back in stock?

Below is a list of common health products out of stock in many cities.

Our algorithm has been searching Amazon for similar products of different presentations and brands.

We can **notify you** if a similar product in our list is in stock and if it can be delivered in 2 weeks or less.

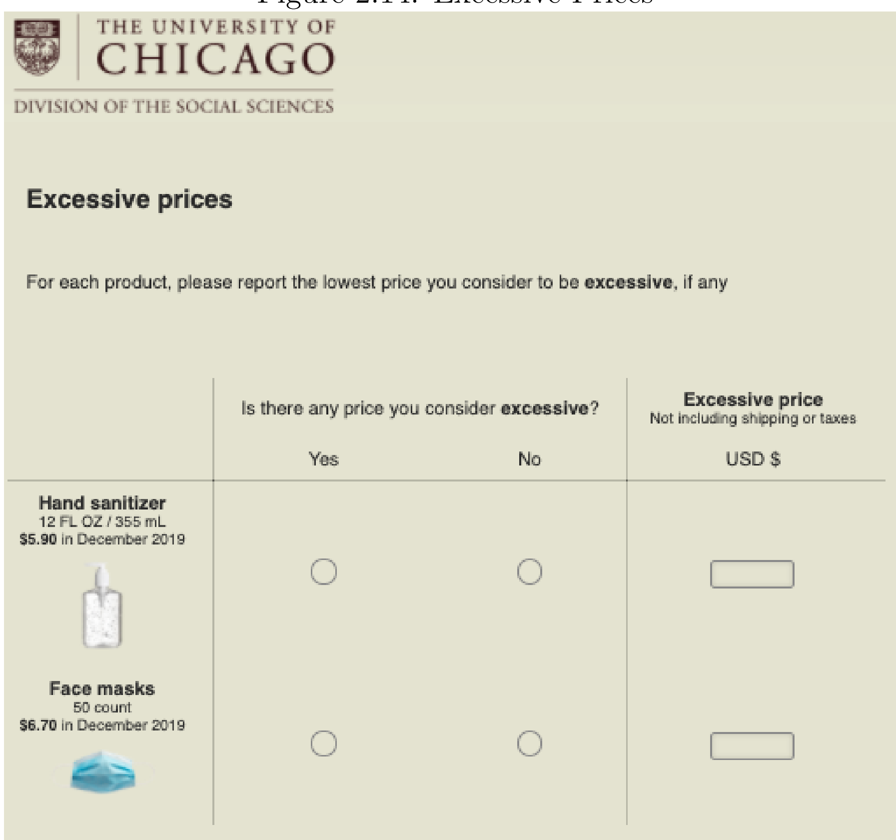
If you want to receive a notification, please enter the **maximum price** that you are willing to pay in the box below. We include average prices of similar products in 2019 as reference.

Prices do not include shipping or taxes

At the end of this survey we will **give you a link** to a randomly chosen product in our list in the price range that you enter (if any).

	Get notified?		**Maximum price** Not including shipping or taxes
	Yes	No	USD $
Hand sanitizer 12 FL OZ / 355 mL **$5.90** in December 2019	○	○	
Face masks 50 count **$6.70** in December 2019	○	○	

Figure 2.14: Excessive Prices



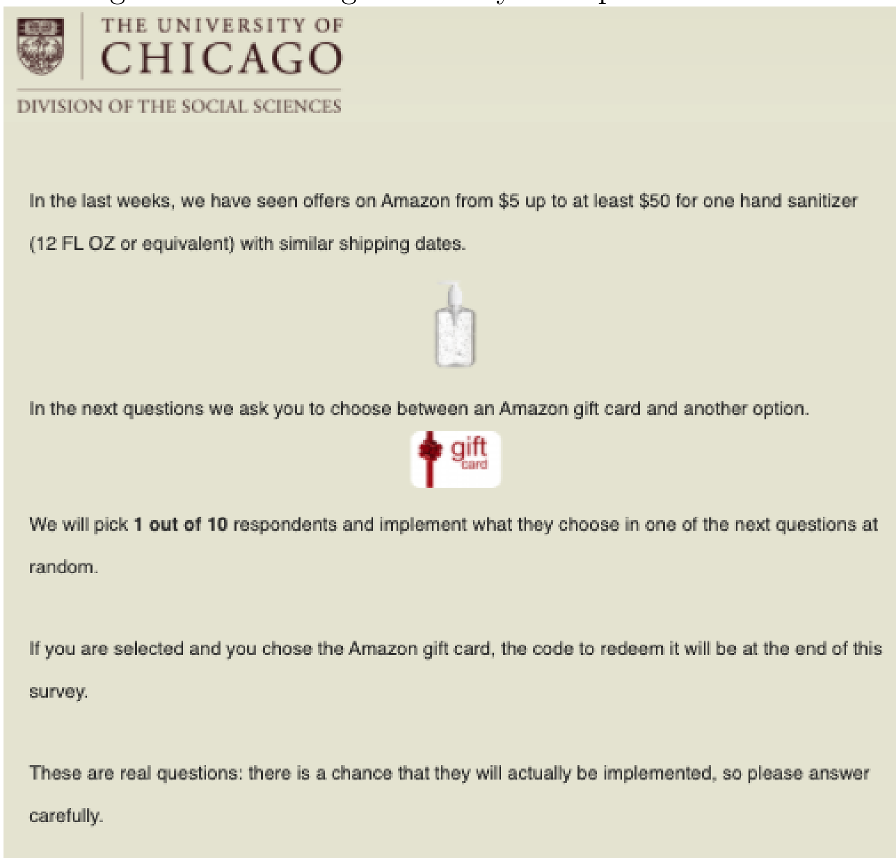
THE UNIVERSITY OF
CHICAGO
DIVISION OF THE SOCIAL SCIENCES

Excessive prices

For each product, please report the lowest price you consider to be **excessive**, if any

	Is there any price you consider **excessive**? Yes	No	**Excessive price** Not including shipping or taxes USD $
Hand sanitizer 12 FL OZ / 355 mL **$5.90** in December 2019	○	○	
Face masks 50 count **$6.70** in December 2019	○	○	

Figure 2.15: Willingness to Pay to Report Instructions


THE UNIVERSITY OF CHICAGO

DIVISION OF THE SOCIAL SCIENCES

In the last weeks, we have seen offers on Amazon from $5 up to at least $50 for one hand sanitizer (12 FL OZ or equivalent) with similar shipping dates.

In the next questions we ask you to choose between an Amazon gift card and another option.

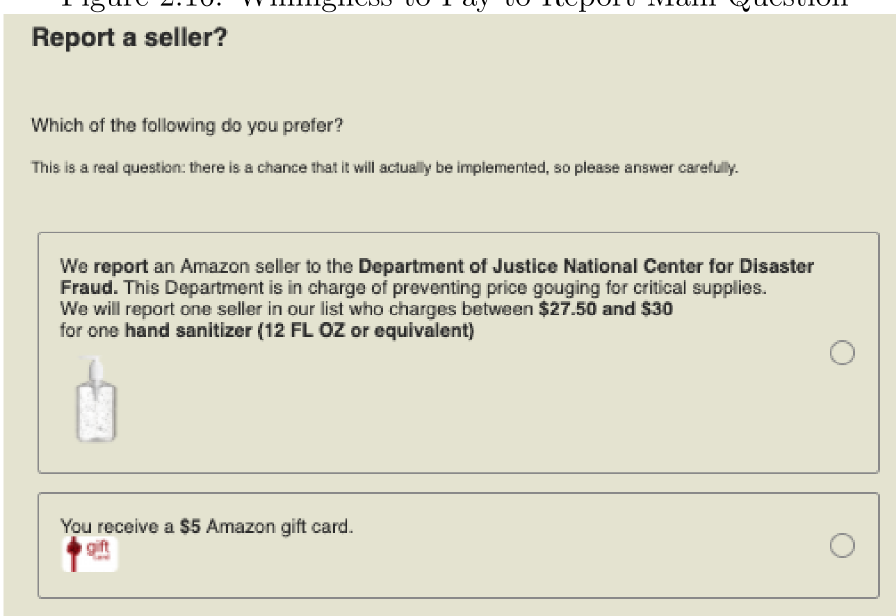
gift card

We will pick **1 out of 10** respondents and implement what they choose in one of the next questions at random.

If you are selected and you chose the Amazon gift card, the code to redeem it will be at the end of this survey.

These are real questions: there is a chance that they will actually be implemented, so please answer carefully.

Figure 2.16: Willingness to Pay to Report Main Question


Report a seller?

Which of the following do you prefer?

This is a real question: there is a chance that it will actually be implemented, so please answer carefully.

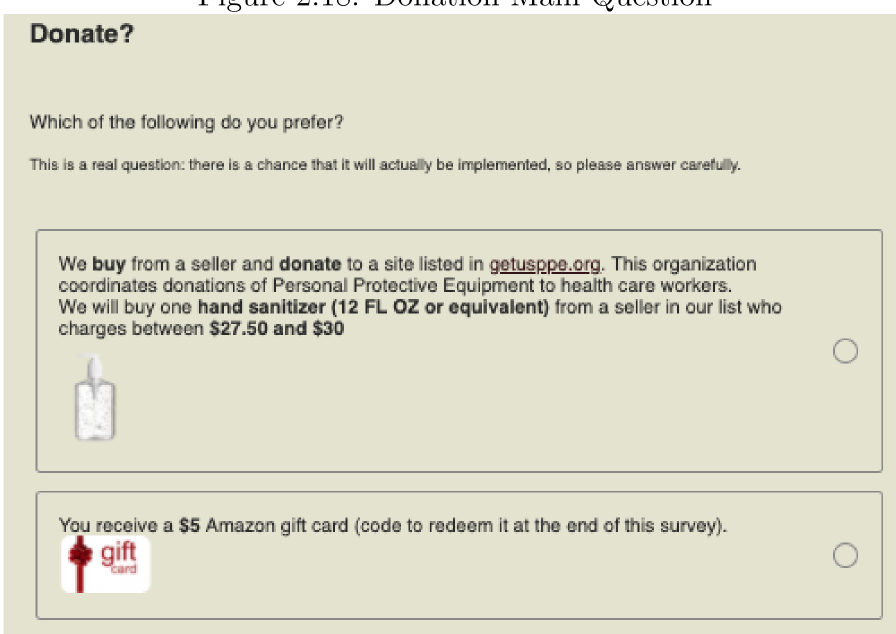
We **report** an Amazon seller to the **Department of Justice National Center for Disaster Fraud.** This Department is in charge of preventing price gouging for critical supplies.
We will report one seller in our list who charges between **$27.50 and $30** for one **hand sanitizer (12 FL OZ or equivalent)**

You receive a **$5** Amazon gift card.

gift

Figure 2.17: Donation Instructions

Figure 2.18: Donation Main Question



CHAPTER 3

THE $100 MILLION NUDGE: INCREASING TAX COMPLIANCE OF BUSINESSES AND THE SELF-EMPLOYED USING A NATURAL FIELD EXPERIMENT

with John A. List, Alejandro Zentner, Marvin Cardoza, and Joaquin Zentner[1]

Abstract

This paper uses large-scale natural field experiments to examine the effectiveness of deterrence nudges on tax compliance in the Dominican Republic. In collaboration with the tax authority, we sent messages to 84,500 businesses who collectively paid $800 million in the year before the experiment. We find that increasing the salience of prison sentences or the public disclosure of evasion increases tax revenue by $193 million (0.23% of GDP). Using a unique sample of large firms, we show that the largest firms, who pay 84% of all corporate income taxes, are considerably more responsive to nudges than typically-studied smaller firms.

1. We would like to thank Chris Blattman, Alec Brandon, Fiona Burlig, Leonardo Bursztyn, Guillermo Cruces, Josh Dean, Dhammika Dharmapala, Steven Durlauf, Julio Elias, Chang-Tai Hsieh, Clement Imbert, Dmitri Koustas, Michael Kremer, Maximiliano Lauletta, Luis Martinez, David Novgorodsky, Nicole Ozminkowski, Ricardo Perez-Truglia, Julia Seither, Cass Sunstein, Dick Thaler, Rebecca Wolfe, Karen Ye as well as seminar participants at University of Chicago Harris, University of Chicago Experiments, Universidad del CEMA the Advances in Field Experiments conference, Behavioral Insights Global, CESifo Workshop on Political Economy, Behavioral Science Summit, MWIEDC, and ZEW Public Finance Conference. The findings, interpretations, and conclusions in this paper are entirely those of the authors and do not represent the views of the Dominican Republic DGII or Inter-American Development Bank. Our study was approved by the University of Chicago Institutional Review Board, IRB19-0510. The tax experiments reported in this paper are registered at the AEA's Social Science Registry, numbers AEARCTR-0004064 and AEARCTR-0004096.

3.1 Introduction

Economic growth critically depends on a country's ability to enforce laws, regulate markets, and provide public goods (Smith, 1776; Kaldor, 1963). The provision of these services depends, in turn, on a state's fiscal capacity. Higher tax evasion rates in developing countries generally lead governments to raise substantially lower tax revenue as a share of GDP than higher-income countries.[2] For example, the 2018 tax to GDP ratio in the Dominican Republic was 13.2% compared to 40% in developed countries. This paper asks whether countries with low enforcement capacity can use nudges to increase tax revenues (Thaler and Sunstein, 2009).

A growing empirical literature documents that governments can use deterrence nudges to increase small- and medium-sized firms' tax compliance (Slemrod et al., 2001; Ariel, 2012; Kleven et al., 2011; Pomeranz, 2015; Slemrod, 2016; Bérgolo et al., 2017). However, taxable income and tax liabilities tend to be highly concentrated in the largest firms. For example, in the Dominican Republic, the top 1% of firms by size pay over 60% of all income taxes. Understanding how nudges affect these firms is essential for the analysis of tax enforcement.

To lend empirical insights into this question, we partner with the internal revenue service in the Dominican Republic (IRSDR hereafter). We sent nearly 84,500 messages to self-employed individuals and firms who collectively paid $800 million in income taxes in FY2018. A unique feature of our field experiment is that we observe firms across the full distribution of size. This feature allows us to test at scale whether these nudges can effectively raise revenues. Furthermore, it allows us to contrast the behavior of small firms to that of large firms, a comparison that to our knowledge has heretofore remained unexplored.

In sum, we randomly allocated subjects into one of six groups. Our control group is a

2. A large literature documents a positive relationship between economic development and the tax to GDP ratio (Rauch and Evans, 2000; Acemoglu, 2005; Besley and Persson, 2009; Acemoglu et al., 2015; Dincecco and Katz, 2016). (Bergeron et al., n.d.) finds that low levels of enforcement place a ceiling on the tax rate for the country.

simple reminder message sent to taxpayers before the tax deadline. Before our partnership, this was standard practice for the IRSDR. To augment this baseline, we include two deterrence messages. One message increased the salience of potential prison penalties instituted under a new law, while the other increased the salience of potential social punishments by emphasizing potential public disclosure of punishments. Then, we interact each message type with a reframing of tax mistakes as voluntary choices on the part of the taxpayer (Hallsworth et al., 2015) to understand how these punishments mediate the treatment effects.

Our key outcome variable is the change in the taxes paid by subjects across the six experimental cells. Guyton et al. (2021) note that an important difficulty in the study of tax evasion by the wealthiest individuals in the United States is that tax evasion can be highly complex and secretive. By sending messages shortly before the tax deadline and comparing the behavior between randomly allocated treatments, we overcome this issue and can interpret any changes in tax payments as changes in evasion.

We report three main results. First, treated subjects collectively pay \$192 million (0.23% of 2018 GDP) more in tax revenue than the control group. Difference-in-differences estimates of the treatment effects suggest that both the threat of public disclosure and prison sentences substantially increase tax compliance. For firms, the public disclosure message increases the amount of taxes paid by an average of \$2,200. In contrast, the prison message increases taxes paid by \$5,300 (we report all monetary figures in USD). [3]

Interestingly, changes in taxable deductions primarily drive our observed treatment effects. We nudge firms that would have taken enough deductions to pay no taxes to report positive taxable income. Self-employed workers also increase taxes paid in response to the deterrence messages. However, the measured effects are smaller than for firms: both message types increase tax revenue from self-employed workers by roughly 12% relative to a simple reminder message. In contrast, the public disclosure and prison messages increase

3. We use the exchange rate as of April 1st, 2019.

tax revenue from firms by 18% and 44%, respectively.

Second, we find that nudge effectiveness is critically moderated by firm size. We find that the largest firms are the most likely to respond to our deterrence messages. For example, our most effective treatment induces the smallest firms to increase their taxes paid by 13 percentage points in 2019. In contrast, the same treatment induces the largest firms to increase their taxes paid by 45 percentage points in 2019. Importantly, nearly $150 million of the total additional tax revenue comes from the largest firms in our sample with the majority of this revenue coming from the prison and prison with commission treatments. This result implies that not having access to the largest firms in the population, as in the previous literature, misses a key feature from a behavioral and public policy perspective.

Third, we find that commission framing effects heavily depend on the perceived underlying punishment and subject pool. When accompanying the control or public disclosure messages, informing subjects that the IRSDR may view mistakes as an active choice negligibly affects tax compliance. However, adding the commission frame to the prison message doubles its magnitude. In this case, firms in our sample increase tax compliance by roughly 100% while the treatment increases the tax compliance of self-employed workers by 21%. The mediating effects of the underlying punishment suggest a possible explanation for the disparate findings in (Hallsworth et al., 2015), who find that commission framing doubles taxes paid, and (De Neve et al., 2019) who find no effect.

We view our findings as having implications for the academic literature and policymakers. In contrast to previous work, we show that the responsiveness to nudges may be increasing in firm size (Kleven et al., 2016; Kumler et al., 2013; Pomeranz, 2015). This evidence supports Guyton et al. (2021)'s claim that evasion requires investment in an evasion technology and that the technology's adoption is likely to be concentrated at the top of the income distribution. From a policy perspective, understanding how the largest firms respond to nudges is invaluable due to the high and increasing concentration of tax liabilities within such firms.

Governments may be leaving large sums of money on the table by only choosing to nudge small- and medium-sized tax payers. Indeed, if we had only sent messages to small- and medium-sized taxpayers, the IRS would have generated $150 million fewer dollars in revenue. In this spirit, our finding that the largest firms are those that respond most strongly to nudges from the tax authority represents a new important insight in the literature and opens up the potential for even broader uses of behavioral nudges in other markets. This result is also important in the context of the current White House's initiative in the US of fighting evasion by high earning corporations and individuals (Tankersley, 2021).

A common consideration with any empirical estimates pertains to generalizability. We use the List (2020) SANS conditions to understand the generalizability of these results. Our sample is a selected subset of the firms and self-employed taxpaying entities in the Dominican Republic. In particular, the IRSDR sent messages to a sample of more than 84,000 self-employed and firms that reported earning more revenue than the average firm in the Dominican Republic in the previous year. In terms of attrition, our compliance rates are 100%, as we have records of the amount of taxes paid for everyone in our sample. Considering the naturalness of the choice task, setting, and time frame, we use a natural field experiment (see Harrison and List (2004)); thus, our setting is one in which subjects are engaged in a natural task and margin. Finally, in terms of scaling our insights, the results suggest that the benefit-cost profile should shrink slightly as the government expands the program to the Dominican Republic's remaining population. This is because we have slightly larger firms in our experiment compared to the overall firm population. Since we view the firm size results as a WAVE1 insight, in the nomenclature of List (2020), replications need to be completed to understand if the size result can be applied to other tax paying populations as well as large players in other markets.

The remainder of our paper proceeds as follows. Section 3.2 describes the institutional context, subjects, and experimental design. Sections 3.3 and 3.4 describe the empirical results. Section 3.5 concludes.

3.2 Background and Experimental Design

In this section, we describe the institutional context of our experiment, the subjects included in the experiment along with their incentives, and the experimental design.

3.2.1 Institutional Context

The Dominican Republic is a Caribbean country with GDP per capita of $8,341 in 2018, and tax revenue was 13% of GDP, much lower than the 22.8% average for all Latin American and Caribbean countries, or the 34.2% average for all OECD countries. The plurality of tax revenue in the Dominican Republic comes from value-added taxes (35.3%). The next largest categories are excise taxes (22.7%), corporate income taxes (16%), individual income taxes (9%), international trade (7.4%) and property taxes (1.4%).

Estimated tax evasion in the Dominican Republic was 61.8% in 2017 (4.22% of GDP) for the corporate income tax and 57.07% for the individual income tax (1.68% of the GDP). This level of tax evasion is higher than in other Latin American and Caribbean countries.[4] To fight high tax evasion, the Dominican Republic's IRS launched an ambitious plan to increase the number of audited taxpayers in 2018. As part of this plan, which involved the matching of taxpayers reported information with third-party information, the audit probability increased from 8% in 2017 to 12% in 2018 (Dirección General de Impuestos Internos, 2020).

While prevalent, tax evasion is seen as a socially undesirable activity in the Dominican Republic. The 1995-1999 World Values Survey shows that 68.6% of the respondents in the Dominican Republic answered that tax evasion is "Never Justifiable," compared to 73% in the United States, 78.3% in Uruguay, 71.7% in Argentina, 52.9% in Mexico, and 46.5% in

4. Tax evasion for the corporate income tax was 39.6% in Uruguay, 49.7% in Argentina, 31.4% in Mexico, and 26.6% in Brazil (Gobierno de la República Dominicana: Equipo Interinstitucional, 2018).

Brazil. More recent information on tax morale is available for Latin American countries from Latinobarómetro. The most recent available wave from year 2016 indicates that 43.6% of the respondents in the Dominican Republic find tax evasion unjustifiable, compared to 58.1% in Uruguay, 62.1% in Argentina, 32.0% in Mexico, and 49.2% in Brazil (Latinobarómetro Corporation, 2016).

In 2018, The Dominican Republic followed regional efforts to combat money laundering and terrorism financing, enacting a law that severely increased the punishment for tax evasion. Law 155-17 against Money Laundering and Terrorist Financing was approved by the parliament in June 2017, regulated by the presidency in November 2017, and further regulated by the Dominican Republic IRS during 2018. It included tax evasion and other tax-related infractions within a list of offenses penalized with severe criminal punishment including prison and stiff monetary fines.[5]

As this law substantially changes the historically low tax enforcement, the IRS commissioner and high-profile political figures have extensively and fiercely discussed its merits in the media. The media has also discussed cases of taxpayers who have been jailed due to tax evasion and fraud. Recently, for example, 21 taxpayers were sentenced to prison or have spent time in preventive detention awaiting trial, 3 have been placed under house arrest, 7 were ordered to use electronic monitoring devices, and 12 have faced travel restrictions. This list includes business owners, managers, and accountants.

In this paper, we focus on corporate and individual income taxes. In the Dominican Republic, employers are responsible for filing the individual income tax for employees, so we only use individual income taxes paid by self-employed individuals. This portion represents 22.2% of the total amount collected by the individual income tax. Table 3.1 shows that the corporate income tax has a flat rate of 27% and the tax rate on self-employed individuals increases with taxable income reaching 27% for the top income bracket. Taxable

5. Because this law was discussed in the media before the tax deadline in 2018; the law might have affected tax compliance in both 2018 and 2019.

income is computed from subtracting expenses and exempted amounts from gross income. In addition to the low-income tax exemption, there are some special regimes providing tax exemptions (e.g., individual educational expenses, firms located in free trade zones, social welfare organizations, charities, and sports). Married couples must file the individual income tax separately.

Table 3.1: Tax Brackets for Firms and Self-Employed Workers

Tax Bracket	Firms	Self-Employed
Less than \$0.00	Exempt	Exempt
\$0.01 - \$8,324.40	27%	Exempt
\$8,324.40 - \$12,486.58	27%	15%
\$12,486.58 - \$17,342.46	27%	\$624.32 + 20% of amount above \$12,486.58
Over \$17,342.46	27%	\$1,595.52 + 25% of amount above \$17,342.46

3.2.2 *Subject Pool*

We conducted our field experiment in collaboration with the IRS of the Dominican Republic (IRSDR). In the months before the experiment, there were 43,973 self-employed workers and 168,497 firms in the agency's database.[6] From this set, the IRSDR sent messages to 28,180 self-employed workers who collectively paid 100 million USD in taxes in FY2018, along with 56,130 firms who collectively paid nearly 700 million USD in FY2018. We randomly split both the self-employed and the firms into six different groups to receive different treatment

6. The database only includes self-employed workers who filed taxes in FY2018. All other self-employed workers are excluded from the experiment and analysis. Not included in our data are the revenues from some micro-sized firms and self-employed workers who opt to use a simplified tax regime based on an estimated minimum tax instead of based on a percentage of the actual income (https://dgii.gov.do/contribuyentesRegistrados/regimenesEspeciales/RST/Paginas/default.aspx), the revenues from one large firm in the mining sector that uses a special tax form to file taxes (form DPUN-01), and the revenues from income taxes collected on capital gains. There are also some differences between the yearly amounts of tax declared and collected, since the IRSDR collects tax debts from previous years.

messages (see Table 3.2).[7]

Table 3.2: Sample Sizes

	Firm	Self-Employed
Control	9,368	4,723
Commission	9,393	4,781
Public Disclosure	9,388	4,736
Public Disclosure with Commission	9,373	4,694
Prison	9,381	4,728
Prison with Commission	9,389	4,518
Total	56,310	28,180

3.2.3 Experimental Design

Each subject's experience follows six steps. First, in FY2018, self-employed individuals and firms decide whether to supply goods and services to the marketplace. Second, at the end of FY2018, the DRIRS sends a message to the subjects with a randomized text. Third, each subject decides whether and when to file their tax return. Fourth, subjects decide how much of their income to report and hence the amount of taxes to pay. Fifth, subjects face the risk of a tax audit by the tax authority. Finally, audited subjects experience the consequences of their decisions while unaudited subjects do not.

As messages were sent shortly before the tax deadline, subjects did not have time to adjust their production decisions before the tax date. Therefore, any changes in the gross-income or losses reported to the DRIRS represent changes in the firms evasion or avoidance decision. In Step 2, the DRIRS sends one of six potential messages to the subject's e-mail

7. Our original intent was to send messages to all 43,973 self-employed workers and 168,497 firms in the IRSDR's sample. However, due to technical issues, the IRSDR sent messages to only the 28,180 self-employed workers and 56,130 firms who we refer to as our experimental sample. Nevertheless, balance was maintained, as treatments in our experimental sample are balanced on observable characteristics.

address and their "virtual office" that almost all taxpayers use to file taxes. These messages were sent three days before the tax deadline for self-employed workers. Firms received the messages twice, thirty and fifteen days before their tax deadline.

Table 3.1 summarizes the natural field experimental design. Our control message is a simple reminder about the tax deadline. Thus, our treatment effects are measured relative to a counterfactual in which the DRIRS reminds subjects about the tax date. The reminder message allows us to control for the potential effect of receiving a letter from the tax authority, which past work has shown in and of itself increases tax compliance (Del Carpio, 2013; Perez-Truglia and Troiano, 2018; Kettle et al., 2016; Mascagni et al., 2017). Accordingly, our measured treatment effects likely understate the effect of the treatment messages relative to an environment in which the tax authority does not contact subjects before the tax deadline.

The experimental variation induced by our five treatments was the inclusion of short phrases after the reminder message. We constructed the phrases to persuade the recipient to correctly report and pay their tax burden by highlighting either the potential for incarceration or public disclosure of punishments. We also interacted both the control message and deterrence messages with an additional paragraph that informed subjects that the tax authority may interpret misreporting as an active choice.

The public disclosure message reminded subjects about the new law and informed them that any punishments levied for tax evasion will be public record available to the population of the Dominican Republic. This message was constructed to vary the subject's perceived probability of their malfeasance and identify concerns for social punishments. This type of punishment may reduce social image utility (Bénabou and Tirole, 2006; Bursztyn and Jensen, 2017; Butera et al., 2019) or raise concerns about losing customers or employees who prefer to work for civically-minded firms (Hanlon and Slemrod, 2009; Du and Vieira, 2012; McDonnell and King, 2013; Servaes and Tamayo, 2013; List and Momeni, 2017; Hedblom et al., 2019). However, since there is no threat of a formal or centralized list meant to shame tax evaders, this message constitutes a weaker form of shaming than countries commonly practice (Hasegawa et al., 2012; Bø et al., 2015; Hoopes et al., 2018; Dwenger and Treber, 2018) or what has been done with experiments targeting individuals (Perez-Truglia and Troiano, 2018).

The second strategy attempted to use the potential deterrence effects of newly implemented prison sentences for tax evasion. To operationalize these treatments, we included a phrase reminding subject's of the newly passed law and that prison sentences are now a potential punishment for tax evasion. This message stands in stark contrast with previous papers on tax compliance which investigate the deterrence effects of financial penalties.[8] Imprisonment, especially of firm owners, is a rare punishment. Given the scarcity of research on the deterrence effects of prison sentences, it is unclear whether this type of deterrence should be viewed as stronger or weaker than traditional deterrence nudges.[9]

We interacted these main messages with a statement that informed subjects that inaccurate information in the tax return may be viewed as an active choice rather than an oversight. These messages are meant to change the behavior of subjects who are evading taxes using an "omission strategy" by increasing the expected punishment for evasion, essentially moving the act of omission to one of commission (Spranca et al., 1991; DeScioli et al., 2011; Hallsworth et al., 2015).

After receiving the messages, subjects proceed to Step 3. Given that we have observed the universe of self-employed workers and firms who choose to file taxes, we have no attrition at this stage of the experiment. However, we can only observe the declarations of self-employed workers and firms who file taxes. Therefore, treatment comparisons of gross and net income

8. See Slemrod (2019) or Antinyan and Asatryan (2019) for reviews of the effect deterrence nudges have on behavior.

9. Lee and McCrary (2017) note that, unlike fines, prison sentences tend to be dispersed across time. This reduces the deterrence effect of prison sentences relative to fines if individuals heavily discount the future or are hyperbolic discounters. However, since firm decisions are made in groups, rather than individually, they may discount future prison sentences less than individuals would (Charness and Sutter, 2012; Denant-Boemont et al., 2017).

Table 3.3: Audit Rates for Firms and Self-Employed Workers

	2018 Either Audit	2018 Automatic Audit	2018 Full Audit	2018 Both Audits
Self-Employed	2360 (8.38%)	2334 (8.28%)	29 (0.10%)	3 (0.01%)
Firms	3781 (6.71%)	3635 (6.46%)	182 (0.32%)	36 (0.06%)
Combined	6141 (7.27%)	5969 (7.06%)	211 (0.25%)	39 (0.05%)

Note: Table displays the number of audits by type and tax-paying entity for all subjects in the experimental sample. These audits occurred before the experiment in FY2018.

can only be made for subjects who chose to file their return. On the other hand, we consider self-employed workers and firms who chose not to file taxes in FY2019 as paying no taxes in that year.

In Step 4, subjects must determine their tax bill. Table 3.1 displays the tax brackets for the subjects in our natural field experiment. Firms face a flat tax of 27% of their taxable income. Self-employed workers, on the other hand, face a progressive marginal tax system. Workers earning less than \$8,324.40 in taxable income are exempt from paying income taxes. Those earning between \$8,324.40 and \$12,468.58 pay 15% of their taxable income over that range. Every dollar earned between \$12,468.58 and \$17,342.46 is taxed at 20% and every dollar earned above \$17,342.46 is taxed at 25%. Both self-employed workers and firms are exempt from paying any taxes if their taxable income is less than or equal to zero.

After making their filing decision, subjects proceed to Step 5, where they face the risk of an audit by the tax authority. Generally, the DRIRS uses two different types of audits. The first is automatic audits that occur when tax entities submit suspicious tax returns. The second type of audit is a full audit where the DRIRS probabilistically chooses some firms to investigate thoroughly. The FY2018 frequency of both types of audits appears in Table 3.3. Historically, there have been low audit rates for both types of subjects. However, this has increased substantially in the past few years.

Finally, the DRIRS may punish audited subjects based on the audit results. Since the enactment of Law 155-17, more than twenty individuals have been sentenced to prison or spent time in pre-trial detention. About two dozen others have been placed under house arrest, ordered to use electronic monitoring devices or faced travel restrictions. This list includes business owners, managers, and accountants. Because the law substantially affected the historical context of low enforcement tax compliance, it has been discussed in the media. The media has also discussed cases of taxpayers who have been sentenced to prison due to tax evasion.

3.3 Results

For precision, we employ a difference-in-differences model as our main specification. In particular, we estimate models of the following form:

$$Y_{i,t} = \alpha + \sum_{g \in G} \gamma_g \mathbb{1}[G = g]_i + \delta_t \mathbb{1}[FY2019]_t + \sum_{g \in G} \beta_g \mathbb{1}[G = g, FY2019]_{i,t} + \epsilon_{i,t}. \tag{3.1}$$

Here, each subject i either receives a reminder message or one of the five treatment messages $g \in G = \{\text{commission, public, public + commission, prison, prison + commission}\}$. There are two time periods, $\mathcal{T} = \{FY2018, FY2019\}$. Outcomes observed in FY2018 are measured pre-treatment and outcomes observed in FY2019 are measured post-treatment. The indicator $\mathbb{1}[FY2019]_t$ is equal to one when the outcomes are after receipt of the treatment messages and the sum $\sum_{g \in G} \gamma_g \mathbb{1}[G = g]_i$ is a set of indicators representing treatment status. Thus, the interaction of each treatment group with the post variable allows us to estimate the causal parameters, β_g, the intent-to-treat effect of receiving a letter with content g relative to the counterfactual of receiving a reminder letter.[10]

10. The identifying assumption needed to interpret β_g as causal effects is that the time trend in outcome Y_{it} for each treatment group would have been the same as the control group had the DRIRS sent the reminder

The primary outcome of interest is the amount of taxes paid. We also consider several intermediate outcomes to learn more about behavioral changes leading to changes in tax evasion. These intermediate outcomes include: whether the subject chose to file, the amounts of gross revenue and net revenue reported conditional on filing, whether the subject chose to declare an exempt-level of income, and whether the subject received an automatic audit for their reporting behavior. We cluster standard errors at the individual level, as this was the level at which we assigned treatments.

3.3.1 Effect of Treatment Messages on Tax Revenue

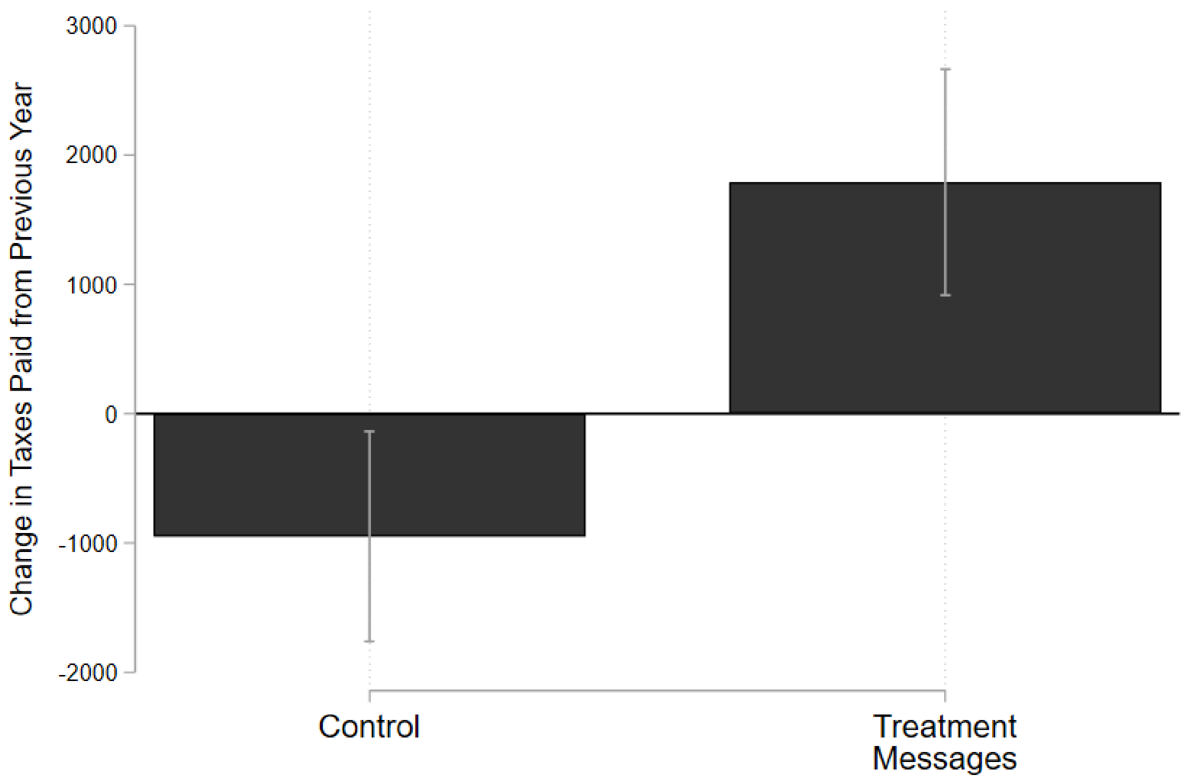
We begin by showing the change in taxes paid from the previous year separately by whether the IRSDR sent a reminder message (the control group) or one of the five nudges for all types of subjects. Figure 3.1 shows that despite the stricter rules against tax evasion, the control group decreases the amount of taxes paid from the previous year by roughly $1,000. This, of course, can be for a variety of reasons that are beyond the scope of this study. Conversely, the pooled treatment group increase the amount of taxes paid by roughly $2,000.

As a falsification exercise, we present an event study analysis of the pooled treatments on taxes paid by firms and self-employed workers. Figure 3.2 displays the difference in taxes paid between treatment and control groups from 2014 to 2019, pooling all treatments and subjects. Under our identifying assumption of statistical independence, taxes paid before treatment assignment should not vary by treatment. We find that there are no pre-treatment differences in taxes paid before we send out the nudges. In the year following the experiment, treatment groups pay $2,700 more on average than the control subjects.

Figure 3.3 presents the estimates of Equation 3.1 using taxes paid as the outcome variable, combining self-employed workers and firms and including a subject-type indicator. Since the

message to those subjects. This condition is satisfied by randomization of the treatments. Estimates of γ_g represent pre-treatment differences in $Y_{i,t}$ between group g and the control. These estimates provide evidence on the validity of the identifying assumption. If these values are not statistically different from zero, then the treatments are well-balanced and we can conclude that the randomization was successful.

Figure 3.1: Average Change in Taxes Paid for Control and Treatment Messages

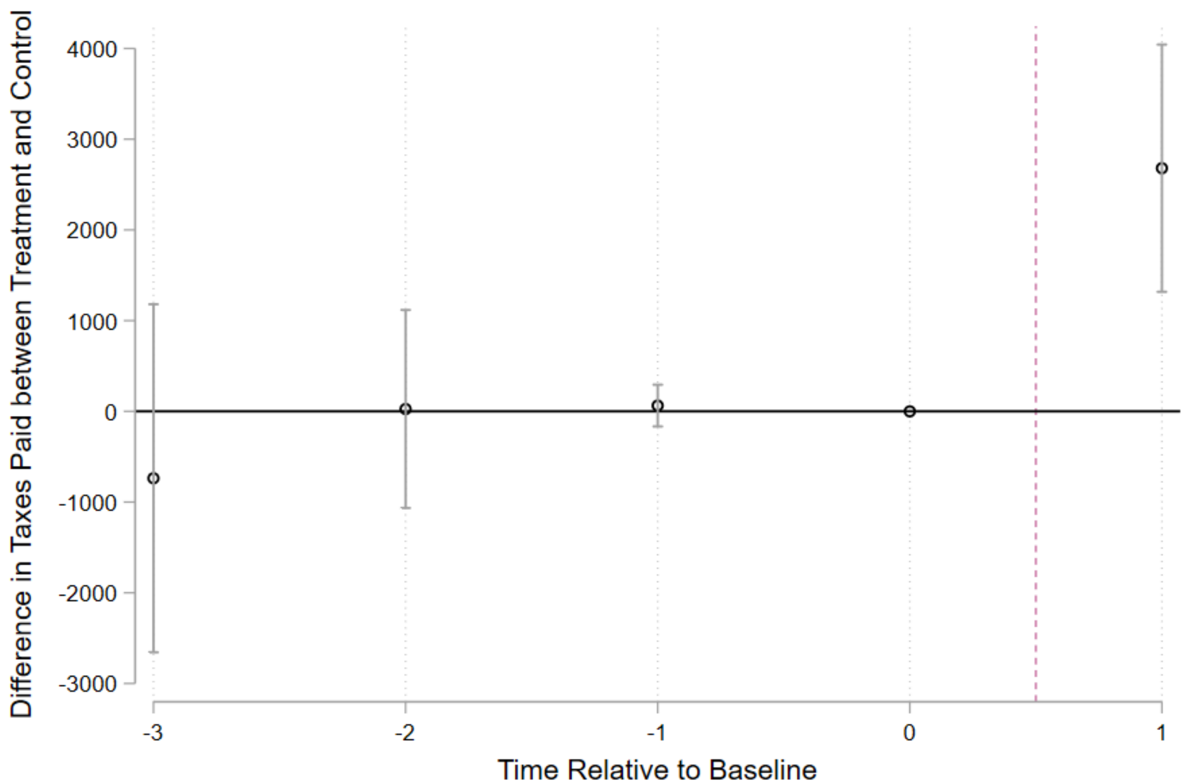


Note: Figure displays the change in tax revenue from FY2018 to FY2019 separately for control and deterrence messages along with 95% confidence intervals. Includes both self-employed worker and firm subjects. The deterrence message group pools the commission, public disclosure, public disclosure with commission, prison and prison with commission treatments.

amount of taxes paid by the control group falls over this time period, estimates of β_g are relative to the control group decline. Overall, both deterrence treatments increase the tax compliance of the subjects. The results show that the public disclosure message increases the amount of taxes paid by $1,616 on average relative to the same change in the control group. The effect for prison messages is about double the public disclosure effect.

Figure 3.4 presents empirical results of this specification using taxes paid as the outcome variable for self-employed workers only, while Figure 3.5 displays the estimated treatment

Figure 3.2: Differences Between Treatment and Control over Time



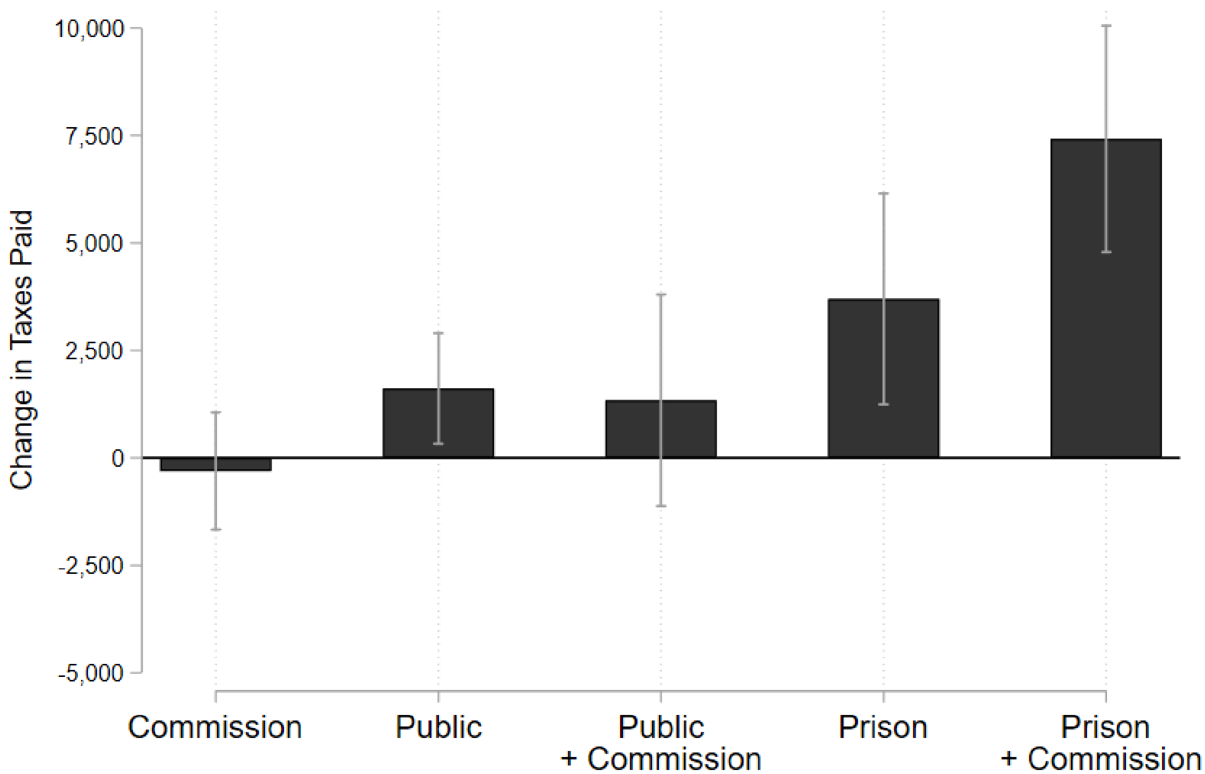
Note: Figure displays the difference in taxes paid between treatment and control groups from 2014 to 2019 with 95% confidence intervals. Includes both self-employed worker and firm subjects. Treatment Effects are calculated using a difference-in-differences model that includes a subject-type indicator. The omitted year is the year used in the randomization, FY2018.

effects for the firm subsample.[11] Both the public disclosure and prison time deterrence messages increase the taxes self-employed workers pay by about \$450, or 13% of the baseline mean. However, the effect of the public disclosure message is imprecisely estimated for self-employed workers. Firms are much more responsive to deterrence messages. Relative to the baseline mean, the public disclosure message increases taxes paid by firms by 19% (\$2,190) and the prison time message increasing taxes paid by 45% (\$5,330).

In aggregate, the control group paid \$13.37 million less in taxes in FY2019 than they

11. Tables including all coefficient estimates appear in the appendix.

Figure 3.3: Average Effects of Treatments on Taxes Paid

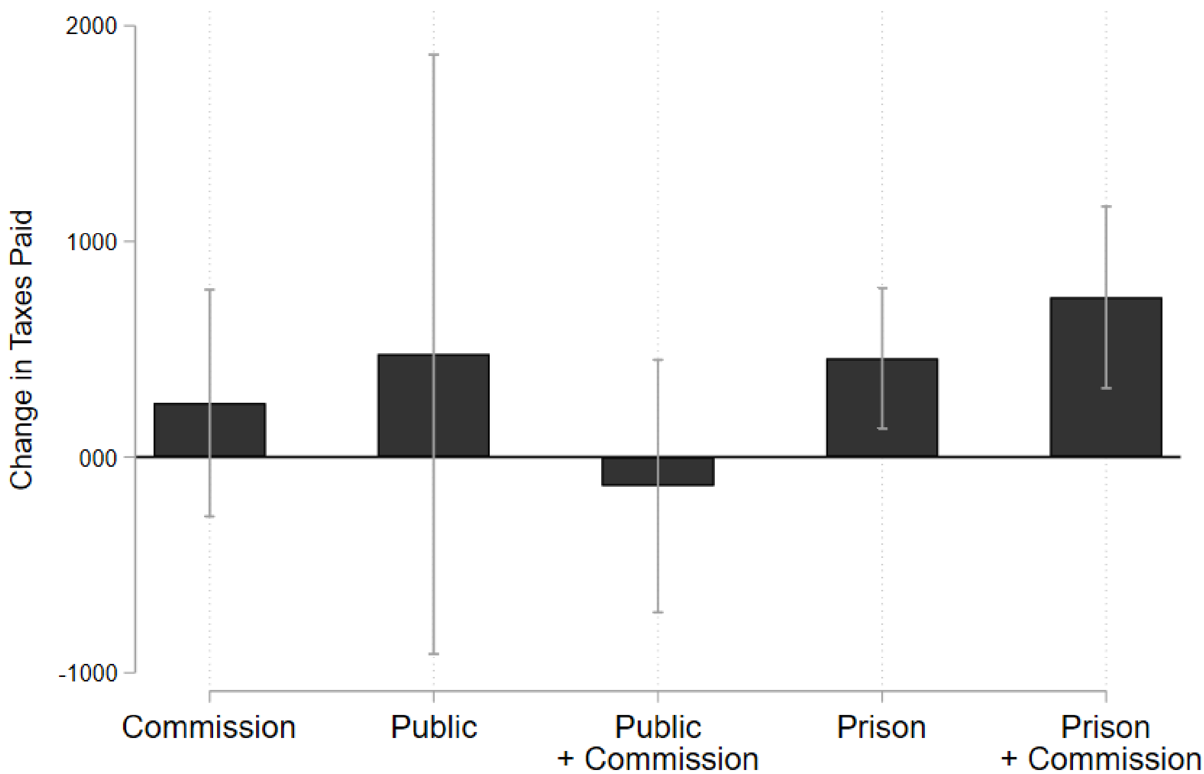


Note: Figures display difference-in-differences coefficients estimated using Equation 3.1, but including a subject type indicator along with 95% confidence intervals. These estimates represent the change in these outcomes due to treatment relative to the control group who received a simple reminder message. Both self-employed workers and firms included. Standard errors are clustered at the unit level.

did in FY2018. Relative to this change, the public disclosure treatment increased taxes paid by \$22.81 million and the prison message increased tax compliance by \$52.17 million (See Figure 3.6). This result highlights the importance of having a control group in FY2019, rather than treating all subjects in FY2019 and using FY2018 as the control (as many tax regulators may desire once they are convinced that the treatment will work).

Potentially at odds with Hallsworth et al. (2015), we find that framing the subject's decision as an active choice did not affect the tax compliance of either self-employed workers or firms. Intuitively, this discrepancy may result from differences in the perceived probability

Figure 3.4: Average Effects of Treatments on Taxes Paid for the Self Employed

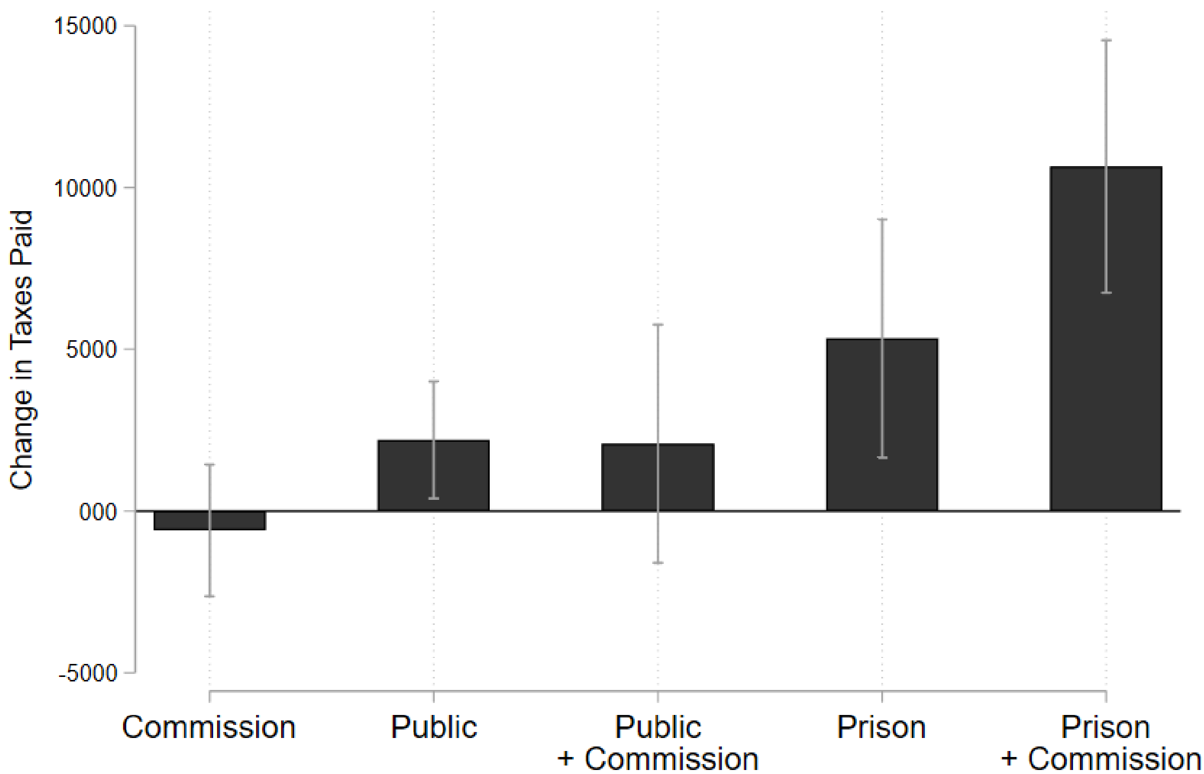


Note: Figures display difference-in-differences coefficients estimated using Equation 3.1 with 95% confidence intervals. These estimates represent the change in these outcomes due to treatment relative to the control group who received a simple reminder message. Only self-employed workers are included. Standard errors are clustered at the self-employed worker level.

of detection or the uncertainty over taxes owed in our setting. In Hallsworth et al. (2015), for example, the tax agency knew the debt owed by each of the subjects. In contrast, the tax burden is not known in either De Neve et al. (2019) or by the DRIRS in our study. Therefore, subjects may not feel unduly threatened by this message.

Figure 3.3 shows that adding the commission frame to the public disclosure message weakly decreases the effectiveness of the deterrent for both self-employed workers and firms. This is likely the result of countervailing forces induced by the treatment. On the one hand, the commission frame increases the perceived harshness of punishments conditional on their

Figure 3.5: Average Effects of Treatments on Taxes Paid for Firms

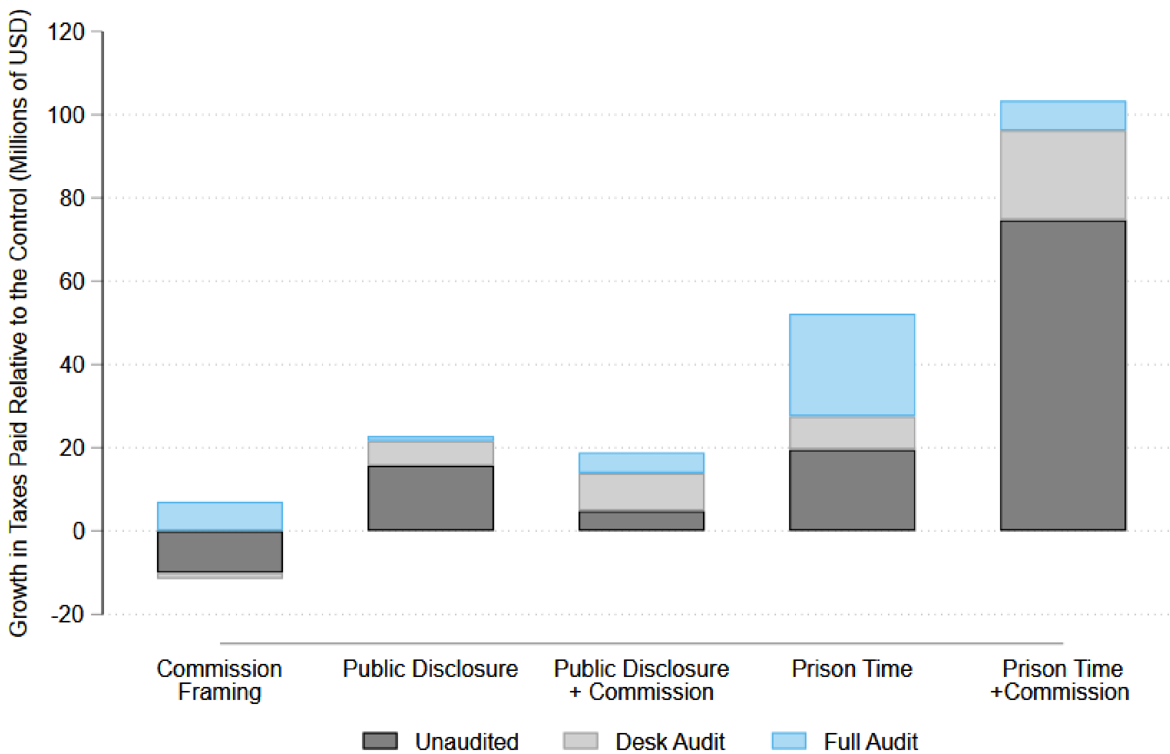


Note: Figures display difference-in-differences coefficients estimated using Equation 3.1 with 95% confidence intervals. These estimates represent the change in these outcomes due to treatment relative to the control group who received a simple reminder message. Only firm subjects are included. Standard errors are clustered at the firm level.

realization. Alternatively, punishing those who make errors dilutes the signal peers receive about the subject's type, conditional on punishment. These countervailing forces could be working against each other in a way that ends up canceling out their effect.

Adding the commission frame to the prison treatment increases the amount of additional taxes self-employed workers pay from 13% to 20% of the baseline mean (see Figure 3.4). Firms respond even more strongly to the addition of the commission frame to the prison message. On average, adding the commission frame to the prison message increases the amount of additional taxes firms pay from 45% of the baseline mean to 90% of the baseline

Figure 3.6: Total Revenue Growth Raised by Treatments



Note: Figure displays the difference between the total tax revenue growth in each treatment less the tax revenue growth in the control group, pooling self-employed workers and firms. Values are calculated separately for each of the three subgroups. These groups are (i) subjects who are never audited, (ii) subjects who received an automatic desk audit, and (iii) subjects who received a full audit. Revenue in the control group fell by $13.37 million to $114.43 million in 2019. Additional $192.79 million raised by treatments.

mean (see Figure 3.5). These effects highlight the sensitivity of the commission language to both the underlying punishment and the sample.

After accounting for the $13 million reduction in taxes from FY2018 to FY2019, adding the commission frame alone decreased total taxes paid by $4.4 million. The effect of adding the commission frame message to the public disclosure message had a similar effect, reducing the effectiveness of the public disclosure message by $3.94 million. In contrast, the prison + commission message increased taxes paid by $103.36 million, $51.19 more than the prison message without the commission frame. In total, the experiment increased tax compliance by \$192.79 million, 0.23% of the Dominican Republic's gross domestic product in 2018.

Next, we estimate the portion of the revenue yield that is new, rather than accelerated. To do this, we use the information on 2019 audits.[12] All of the debt owed by subject's receiving full audits is likely to be recovered by the DRIRS. Therefore, the increase in taxes generated from this group represents accelerated, rather than new revenue. Subjects who receive an automatic audit from the DRIRS will likely have some of their debt recovered. Subjects who are never audited by the DRIRS are unlikely to end up making additional payments to the DRIRS.[13] Thus, we measure the additional yield generated from treatments as the yield from subjects who are never audited and the accelerated yield generated from treatments as the yield from subjects who receive either type of audit.

Figure 3.6 shows the total change in revenue for each treatment relative to the control group. Overall 54% (\$104.8 million) of the tax revenue raised by the treatments constitutes new revenue. Up to an additional \$43.02 million from firms receiving an automatic audit is also new revenue. The \$44.97 million raised by firms who will receive a full audit represents revenue the Dominican Republic would have received absent treatments.[14]

3.4 Firm Size Heterogeneity

A common feature of tax systems is the disproportionate fraction of total revenue paid by the largest firms. For instance, Figure 3.7 shows that in FY2018, the largest twenty percent

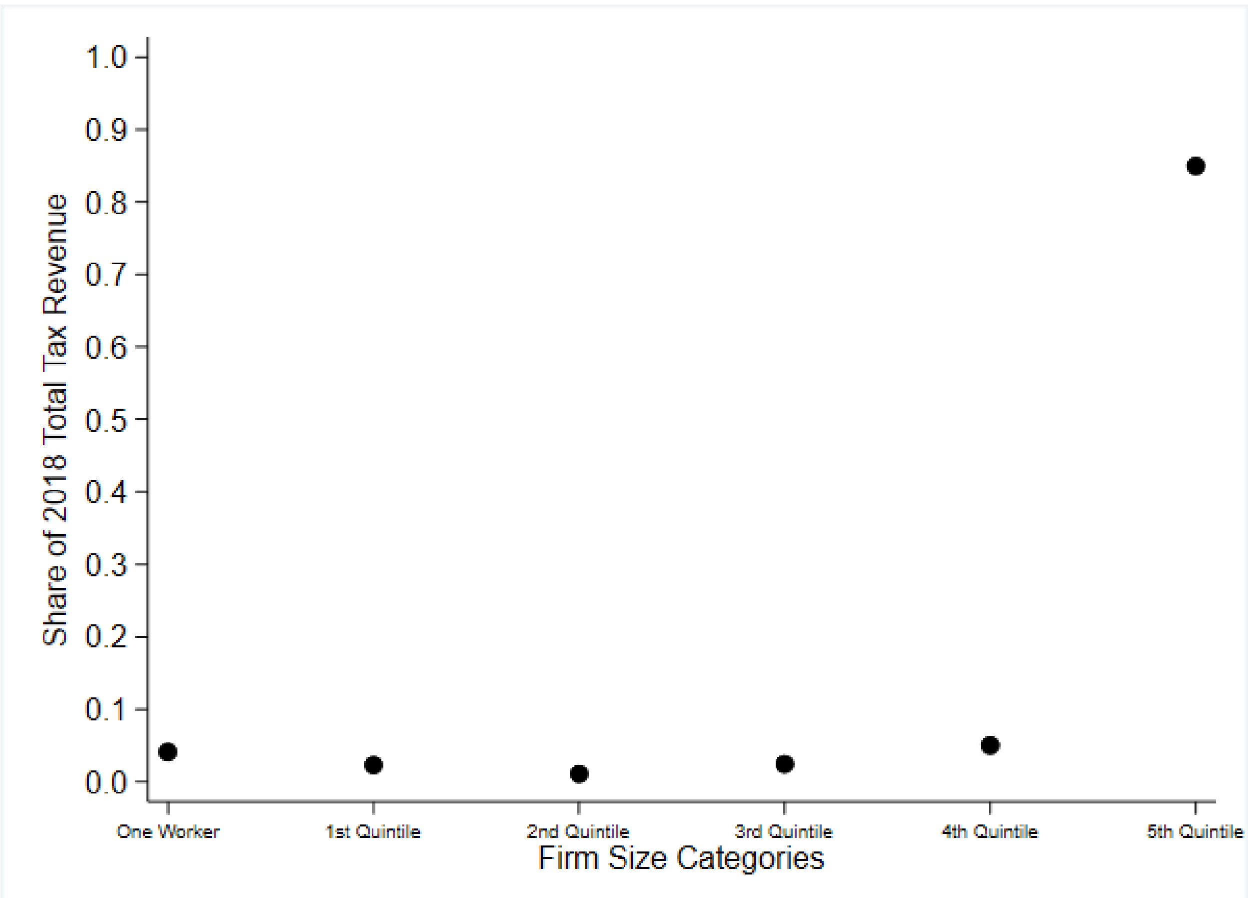
12. We do not have information on the exact date of the audits. So, we do not know for certain whether these numbers reflect pre- or post-audit values. However, almost all of the gains in revenue occur in the month after the experiment and the treatment effects are not affected when we measured them in different months following the intervention, suggesting that the treatment effects would not be influenced by the audits or their timing.

13. The treatment messages may also change the composition of audited firms. Indeed, we do find that self-employed workers in the commission, public, and prison with commission treatments are less likely to receive automatic audit than those in the control group.

14. The calculation of new and accelerated revenue relies on the assumption that there will be no future audits that bring in revenue. If new audits occur, the total amount of money raised by the treatments may fall.

of firms in the Dominican Republic paid over eighty-five percent of all corporate income taxes. The top one percent of firms alone pay over sixty-percent of all income tax revenue in the country. Despite the importance of the largest firms, previous experimental research on tax compliance has focused on interventions with small- and medium-sized firms (see, e.g., Ariel (2012), Pomeranz (2015) and Bérgolo et al. (2017)).

Figure 3.7: Share of Taxes Paid by Firm Size



Note: Figure displays the share of tax revenue paid by firms in each of the firm size categories. Firm size quintile is calculated as the quintile of firm size conditional on having more than one worker. This value is calculated using the size of all firms in the country that pay income taxes.

It is not obvious whether the largest firms will respond to deterrence nudges in the same way smaller and more well-studied firms have responded. Kleven et al. (2016) argue that tax evasion is more difficult in large firms with accurate business records as the threat of whistleblowing increases with the number of employees. In line with this argument, Kumler et al. (2013) find that tax compliance in Mexico is higher in larger firms. Similarly, Pomeranz (2015) finds that the measured effects of threatening random audits are decreasing in firm size.

Alternatively, Guyton et al. (2021) argue that high fixed costs to tax evasion may lead to larger firms having higher levels of evasion. Absent treatment, these firms would have a greater ability to respond to our nudges. Moreover, Pomeranz (2015) notes that a firm owner's priors about audit probabilities, risk-aversion, and other characteristics may vary with firm size. Finally, firms in the Dominican Republic primarily sell directly to consumers. Kumler et al. (2013) and Pomeranz (2015) note that firms with a large share of sales going to final consumers will have an easier time evading taxes. So, the previous results from Pomeranz (2015) may not apply in this setting.

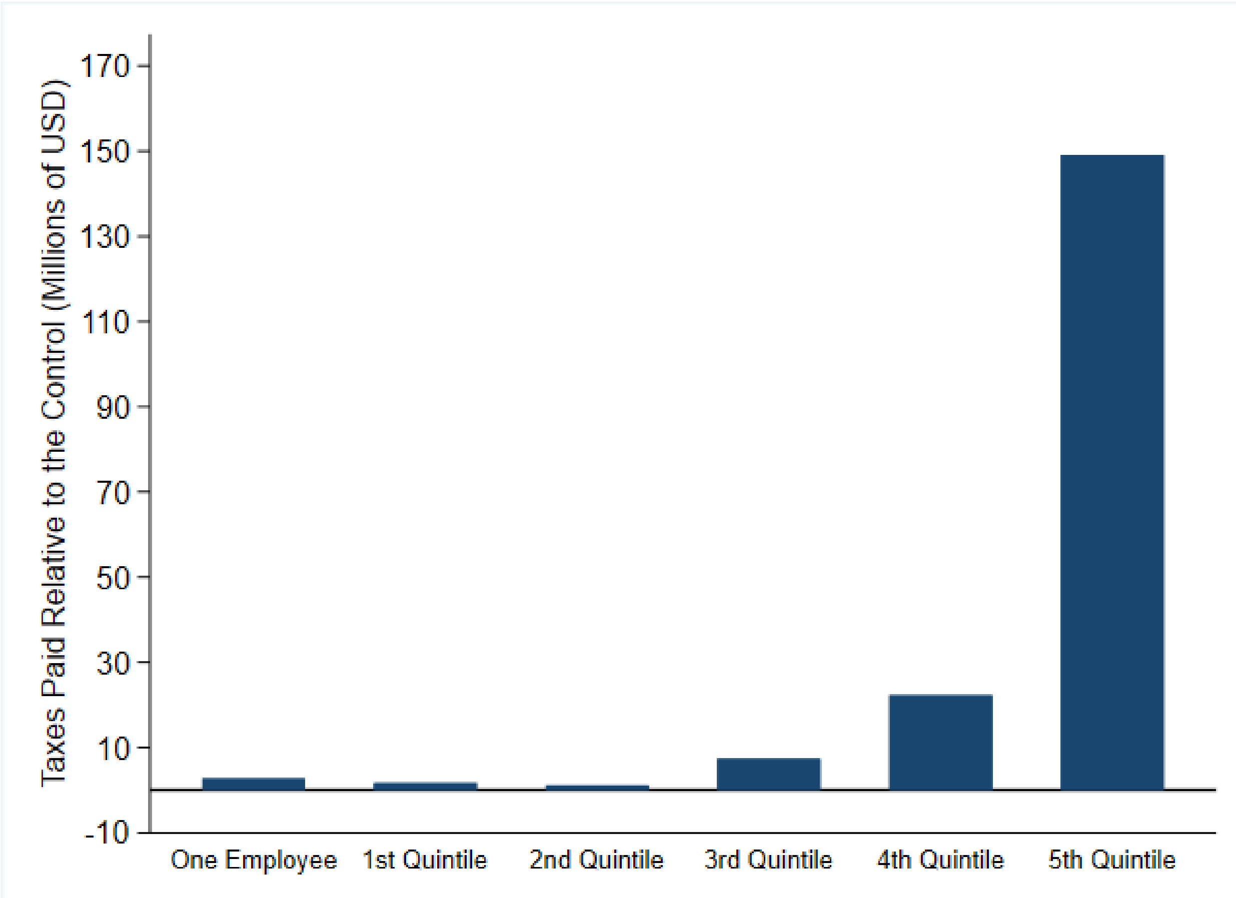
To understand the potential heterogeneous effects of nudge treatments across the distribution of FY2018 firm sizes, we separate firms into six groups based on their number of workers. The first group is firms with a single worker (67% of all firms have a single worker). Then, we separate the data into quintiles conditional on having more than one worker.[15] The discreteness of the data and the skewed distribution means that quintiles do not have exactly 20% of the observations.

Figure 3.8 shows that just under \$150 million tax dollars out of the \$192 million raised by the experiment came from the top group of firms. The next highest group of firms paid around \$20 million additional taxes. The rest came from the bottom three quintiles of firms, firms with one employee and self-employed workers.

Figure 3.9 displays the share of taxes paid by each treatment within the top group of firms and the other subjects. This figure shows that had we experimented purely on the types of firms considered in previous papers, we would have falsely concluded that governments are

15. We create quintiles using the full sample of data, before sub-setting the data into the experimental sample. We break the data into quintiles for statistical power.

Figure 3.8: Total Raised in each Quintile

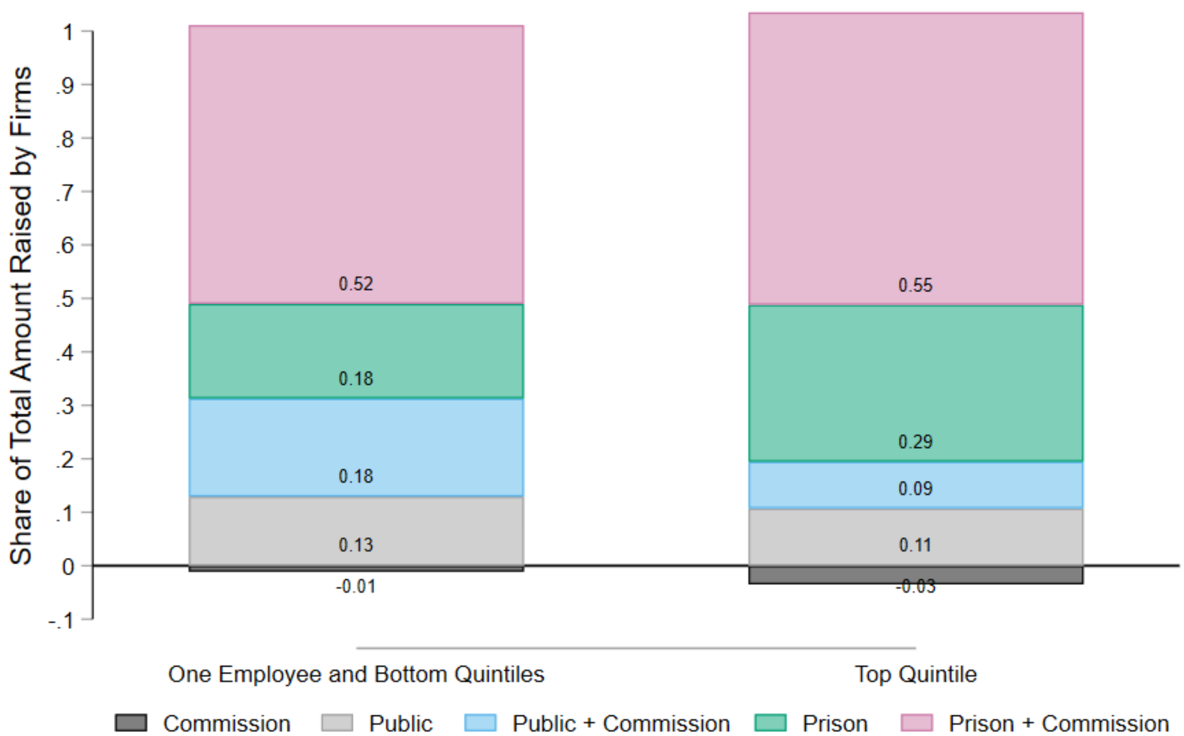


Note: Figure displays the total taxes paid in each quintile over the amount raised by the control group in that firm size category.

just as well off focusing on the public with commission nudges as prison deterrence nudges. However, the prison message raises three times as much revenue as a share of total revenue as the public with commission nudge in the top group of firms.

Next, we examine the treatment effects on the probability of paying more taxes in FY2019 than in FY2018. Figure 3.10 displays the probability that firms in each quintile pay strictly more taxes in FY2019 than FY2018 by treatment and quintile of firm size. We find that firms in the first quintile of firm size are more likely to respond to treatment than firms with only one worker.

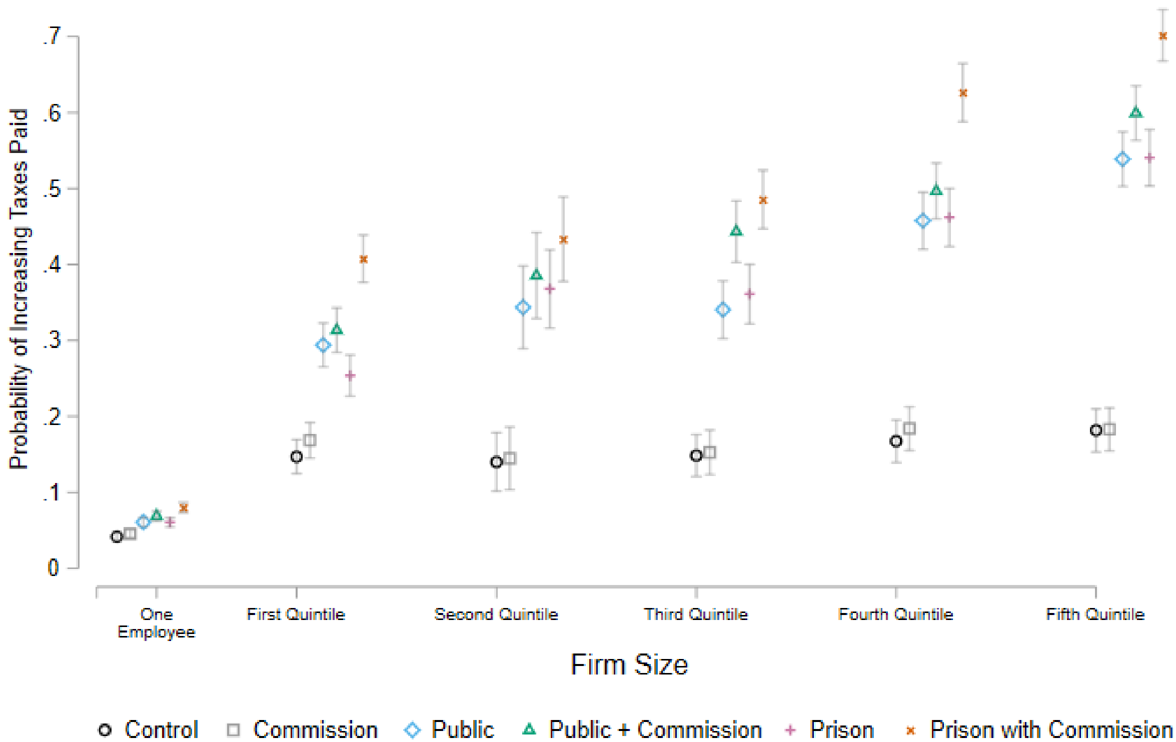
Figure 3.9: Share of Total Raised by Treatment



Note: Figure displays the share of total revenue earned by each treatment separatley for the bottom 80% of firms and top 20% of firms. Firm size categories are firms with only a single employee and then quintiles of firm size conditional on having more than one employee.

However, conditional on having more than one worker, the probability of increasing taxes paid in the control group does not vary with firm size. The commission frame message does not increase the probability of paying more taxes in 2019 at any firm size. However, firms in each of the four deterrence message treatments are more likely to pay more taxes in 2019 at every level of firm size. The size of the treatment effect is increasing with the size of the firm. At the highest quintile, which represents the top 6.5% biggest firms because firms with one employee are excluded when defining quintiles, the prison with commission treatment is the most likely to induce firms to pay more taxes. Next, we present the change in taxes paid over this period in Figure 3.11. Unsurprisingly, we find that the amount of taxes paid

Figure 3.10: Probability of Increasing Taxes Paid by Treatment and Firm Size



Note: Figures displays the probability of paying strictly more taxes in 2019 than 2018 for each treatment group and the control group in a given category. Categories are firms with a single worker and then quintile of firm size conditional on having more than one worker at the firm. Also presented are 95% confidence intervals. The firm size quintile is calculated using the 2018 distribution of the number of employees from the full sample of firms.

by the highest quintile of firm size dwarfs the amount of taxes paid by lower quintiles in importance.

A key result is that we find that the top 20% of firms have the largest treatment effect. The average taxes paid by control firms for this quintile fell by \$13,643. During this same period, taxes paid by the public disclosure group increased by \$9,171 when used alone and \$3,324 when combined with the commission treatment. The prison deterrence messages were much more effective also, increasing taxes paid by \$42,299 when used alone and \$95,846 when

Figure 3.11: Change in Taxes Paid by Treatment and Firm Size

Note: Figures displays change in taxes paid between FY2018 and FY2019 for each of the treatment groups and the control group in a given category. Categories are firms with a single worker and then quintile of firm size conditional on having more than one worker at the firm. Also presented are 95% confidence intervals. The firm size quintile is calculated using the 2018 distribution of the number of employees from the full sample of firms.

combined with commission. These effects contrast with Pomeranz (2015) and Kumler et al. (2013) who find that tax compliance is lower in smaller firms. This discrepancy may be due to differences in the paper trail associated with the corporate income taxes compared to VATs, or the large portion of firms in our study who sell to final consumers.

The different results in our setting can also be explained by the nature of our intervention compared to the setting in the theoretical model in Kleven et al. (2016) and in the empirical evidence provided by Pomeranz (2015) and Kumler et al. (2013). Specifically, large firms might consider prison as a more likely consequence of tax evasion than small firms, implying that larger firms might have found the prison message as a more credible threat than smaller firms. A similar mechanism might be argued for the threat of public disclosure. More work studying heterogeneous effects by firm size for different types of interventions is necessary.

In a general sense, we view our firm size results as novel in the sense that we are first experiment showing that the biggest players in a market are most deeply affected by behavioral nudges. Since the modal research insights from behavioral and experimental studies are derived from individuals and smaller players in a market, our results open up a potential avenue for future research using behavioral interventions on market players that can impact prices and allocations at the most fundamental level.

3.5 Conclusions

Understanding the sources and underpinnings of economy-wide growth remains a key fixture amongst economists and policymakers alike. Acemoglu et al. (2015) emphasize the importance of state capacity which depends directly on a country's fiscal capacity (Kaldor, 1963). The low levels of compliance in developing countries can lead to negative feedback loops from taxation to development (Besley and Persson, 2014). Such structural features suggest that the complementarity between taxation and development represents a key to unlocking economic growth in poor countries.

Over the past decade, there has been a surge of literature attempting to increase tax compliance and understand the determinants of tax evasion. We take the literature in a new direction by partnering with the Dominican Republic and experimentally evaluating how prison sentences and non-pecuniary motives influence tax compliance across firms of various sizes (Luttmer and Singhal, 2014). In doing so, we present the first evidence from a natural field experiment testing whether increasing the salience of public disclosure of penalties affects firms' tax compliance behavior. Moreover, in contrast to earlier experiments on tax compliance, we leverage administrative data to shed insights into how behavioral nudges affect firms across the entire size distribution.

A key result from our field experiment is that messages highlighting the potential for incarceration for tax evasion are highly effective in increasing tax compliance. Increasing the salience of potential incarceration increases the amount of taxes paid by 45% for firms and 13% for self-employed workers. Messages that increase the salience of public disclosure are less effective. Such messages do not significantly increase the amount of taxes self-employed workers pay, but increase the amount of taxes firms pay by 19%. Importantly, the channel for the observed effects is primarily driven by dramatic reductions in the propensity of firm's to declare exempt levels of taxable income.

We find that the effect of commission messaging depends greatly on the perceived punishment regime. The framing effect greatly augmented the prison time message, while weakly backfiring when used with a reminder or when increasing the salience of public disclosure. Overall, the treatments increased the amount of income taxes paid by $193 million, with a little over half of this revenue estimated as "new," rather than accelerated revenue. That is, the government would not have raised this revenue without our treatment messages.

A final insight from our data is that nudges are quite effective in driving behavioral change within the largest firms in our sample. This is important from a practical public policy fiscal capacity perspective, since large firms bear a disproportionate amount of the tax burden. While this result is consistent in our data set, we know less about why this result holds. More specifically, are larger firms responding to reputational concerns from customers, or do larger firms exhibit greater social image concerns than smaller firms? Are there certain institutional facts that cause this asymmetry in treatment effects? We trust that future research will replicate our insights and build theoretical frameworks to enhance our understanding of tax compliance across the firm size dimension.

3.6 Appendix

3.6.1 Supplementary Evidence

Table 3.4: Self Employed Baseline Outcome Balance

	(1) Filed Taxes	(2) Gross Revenue	(3) Taxable Revenue	(4) Declared Exempt	(5) Taxes Paid	(6) Auto Audit	(7) Full Audit
FY2018: Commission Frame	0	3541.3	307.6	0.00332	85.17	-0.00669	-0.000434
	(.)	(7409.7)	(1578.0)	(0.00984)	(392.9)	(0.00565)	(0.000666)
FY2018: Public	0	-5956.3	323.7	0.00606	87.31	-0.00889	-0.000426
	(.)	(6623.3)	(2358.6)	(0.00985)	(588.6)	(0.00563)	(0.000669)
FY2018: Public & Commission Frame	0	-7322.9	126.2	0.00651	42.90	-0.00458	-0.000631
	(.)	(7764.5)	(1920.7)	(0.00987)	(478.9)	(0.00571)	(0.000636)
FY2018: Prison	0	-1315.9	-1095.3	0.000592	-268.7	0.00202	0.000422
	(.)	(6894.1)	(1472.8)	(0.00988)	(366.5)	(0.00580)	(0.000791)
FY2018: Prison & Commission Frame	0	8114.3	1149.5	0.00195	295.6	-0.000507	-0.000385
	(.)	(9173.0)	(1547.6)	(0.00999)	(385.0)	(0.00583)	(0.000682)
Constant	1	76550.7***	16210.6***	0.640***	3561.3***	0.0860***	0.00127**
	(.)	(4741.9)	(1124.2)	(0.00699)	(279.9)	(0.00408)	(0.000518)
Observations	28179	28179	28179	28179	28179	28179	28179

Note: Table displays regressions of treatment indicators on pre-treatment outcomes with robust standard errors in parentheses. Only includes self-employed workers who received treatment messages from the IRSDR. Note that every subject in our self-employed sample filed taxes in 2018. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.5: Firm Baseline Outcome Balance

	(1) Filed Taxes	(2) Gross Revenue	(3) Taxable Revenue	(4) Portion Taxable	(5) Declared Exempt	(6) Auto Audit
FY2018: Commission Frame	0.000337	127211.4	5676.9	0.00265	-0.0162	0.0123
	(0.00726)	(259341.9)	(29167.9)	(0.00672)	(0.0112)	(0.0127)
FY2018: Public	0.000203	30489.4	8701.3	-0.000491	-0.0151	0.000980
	(0.00726)	(249369.6)	(28817.4)	(0.00671)	(0.0110)	(0.0119)
FY2018: Public & Commission Frame	-0.000840	259399.9	-11900.5	0.00288	-0.0133	0.00993
	(0.00727)	(379504.7)	(38984.0)	(0.00673)	(0.0106)	(0.0122)
FY2018: Prison	-0.000241	168222.9	25935.6	0.00144	0.00344	-0.00204
	(0.00727)	(367515.5)	(42529.1)	(0.00672)	(0.00948)	(0.0115)
FY2018: Prison & Commission Frame	0.000357	13688.6	11041.0	0.00299	-0.00371	-0.00430
	(0.00726)	(263628.3)	(32687.7)	(0.00672)	(0.00997)	(0.0113)
Constant	0.548^{***}	1122403.6^{***}	55424.5^{**}	0.304^{***}	0.969^{***}	0.0455^{***}
	(0.00514)	(185847.5)	(23379.4)	(0.00475)	(0.00690)	(0.00825)
Observations	56310	30877	30877	56310	3781	3781

Note: Table displays regressions of treatment indicators on pre-treatment outcomes with robust standard errors in parentheses. Only includes firms who received treatment messages from the IRSDR. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.6: Firm Representativeness

	(1) Employees	(2) Filed Taxes	(3) Gross Revenue	(4) Taxable Revenue	(5) Declared Exempt	(6) Taxes Paid	(7) Auto Audit	(8) Full Audit
Included in Experiment	1.288^{***}	-0.00442^{*}	650228.9^{***}	44635.4^{***}	-0.0673^{***}	7448.7^{***}	-0.0257^{***}	0.0295^{***}
	(0.463)	(0.00257)	(103815.0)	(12085.6)	(0.00242)	(1660.4)	(0.00530)	(0.00620)
Constant	8.228^{***}	0.553^{***}	571907.0^{***}	17371.4^{***}	0.373^{***}	4828.7^{***}	0.987^{***}	0.0187^{***}
	(0.273)	(0.00148)	(36778.8)	(6172.6)	(0.00144)	(857.4)	(0.00428)	(0.00513)
Observations	168106	168497	92889	92889	168497	168497	4478	4478

Note: Table displays regression coefficeints from regressions of FY2018 baseline outcomes on an indicator for whether the firm was sent a message. Robust standard errors shown in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.7: Self-Employed Representativeness

	(1) Filed Taxes	(2) Gross Revenue	(3) Taxable Revenue	(4) Declared Exempt	(5) Taxes Paid	(6) Auto Audit	(7) Full Audit
Included in Experiment	0	-1966.7	5674.8^{***}	-0.132^{***}	1274.8^{***}	0.00957^{***}	0.000206
	(.)	(4131.4)	(684.5)	(0.00438)	(170.3)	(0.00264)	(0.000298)
Constant	1	77977.2^{***}	10663.9^{***}	0.775^{***}	2325.0^{***}	0.0733^{***}	0.000823^{***}
	(.)	(3400.2)	(393.1)	(0.00332)	(97.41)	(0.00207)	(0.000228)
Observations	43972	43972	43972	43972	43972	43972	43972

Note: Table displays regression coefficients from regressions of FY2018 baseline outcomes on an indicator for whether the self-employed worker was sent a message. Robust standard errors shown in parentheses. All self-employed workers in our sample filed taxes in 2018, so there are no differences between units included and excluded from our experimental sample. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.8: Average Effects of Treatment Messages on Change in Taxes Paid

	(1) Taxes Paid	(2) Taxes Paid	(3) Taxes Paid
Commission Framing	-307.9	250.8	-595.8
	(828.5)	(319.8)	(1237.9)
Public Disclosure	1616.1**	477.0	2190.0**
	(783.7)	(844.3)	(1099.0)
Public Disclosure with Commission Framing	1338.4	-134.0	2077.3
	(1494.1)	(355.6)	(2235.6)
Prison Time	3697.7**	457.6**	5330.2**
	(1490.0)	(198.1)	(2238.5)
Prison Time with Commission Framing	7419.1***	740.4***	10645.5***
	(1602.2)	(255.7)	(2372.5)
2019 indicator	-947.8*	-347.7**	-1249.8*
	(493.4)	(166.0)	(736.9)
Pre-treatment: Commission Framing	174.5	85.17	214.5
	(2545.6)	(392.9)	(3828.1)
Pre-treatment: Public Disclosure	246.0	87.31	324.9
	(2505.4)	(588.6)	(3755.9)
Pre-treatment: Public Disclosure with Commission Framing	-23.89	42.90	-55.05
	(3151.1)	(478.9)	(4726.1)
Pre-treatment: Prison Time	1466.3	-268.7	2339.9
	(3999.1)	(366.5)	(6010.9)
Pre-treatment: Prison Time with Commission Framing	9.437	295.6	-107.1
	(3001.5)	(385.0)	(4472.4)
Constant	2607.3	3561.3***	11824.7***
	(1875.6)	(279.9)	(2932.9)
Subjects	Self-Employed and Firms	Self-Employed	Firms
Subject Indicator	YES	NO	NO
Baseline balance	.997	.618	.996
Observations	168978	56358	112620

Note: Table displays coefficient estimates of Equation 3.1 using amount of taxes paid as the outcome variable. Standard errors are clustered on the individual level. Baseline Balance row displays the p-value from an F-test evaluating whether all pre-treatment indicators are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.9: Propensity to Declare Exempt Income Level

	(1) Exempt Net-Income Level	(2) Exempt Net-Income Level	(3) Exempt Net-Income Level	(4) Exempt Net-Income Level
Commission	-0.0189**	-0.0191**	-0.0371***	-0.0549***
	(0.00931)	(0.00943)	(0.0106)	(0.0140)
Public	-0.0360***	-0.0371***	-0.180***	-0.260***
	(0.00954)	(0.00968)	(0.0103)	(0.0134)
Public & Commission Frame	-0.0415***	-0.0428***	-0.213***	-0.302***
	(0.00966)	(0.00978)	(0.0103)	(0.0134)
Prison	-0.0479***	-0.0488***	-0.185***	-0.272***
	(0.00973)	(0.00985)	(0.0103)	(0.0135)
Prison & Commission Frame	-0.0664***	-0.0679***	-0.234***	-0.341***
	(0.00970)	(0.00983)	(0.0103)	(0.0134)
2019 indicator	0.00176	-0.00887	0.153***	0.255***
	(0.00656)	(0.00666)	(0.00761)	(0.00996)
FY2018: Commission Framing	0.00332	0.00386	0.00449	0.000910
	(0.00984)	(0.00984)	(0.00979)	(0.0114)
FY2018: Public	0.00606	0.00613	-0.00110	-0.00907
	(0.00985)	(0.00986)	(0.00980)	(0.0114)
FY2018: Public & Commission Frame	0.00651	0.00725	0.00610	-0.00410
	(0.00987)	(0.00988)	(0.00980)	(0.0114)
FY2018: Prison	0.000592	0.000745	0.00287	0.00185
	(0.00988)	(0.00988)	(0.00980)	(0.0114)
FY2018: Prison & Commission Frame	0.00195	0.00228	0.00509	0.00720
	(0.00999)	(0.00999)	(0.00979)	(0.0114)
Constant	0.640***	0.640***	0.554***	0.339***
	(0.00699)	(0.00699)	(0.00693)	(0.00804)
Subjects	Self-Employed	Self-Employed	Firms	Firms
Gross Income Declared	Any	Positive	Any	Positive
Baseline balance	0.970	0.965	0.956	0.67
Observations	50703	50035	59331	40623

Note:Table displays coefficient estimates of Equation 3.1 using propensity to declare an exempt level of taxable income as the outcome variable. Only subjects who filed taxes are included. Standard errors are clustered on the individual level. Baseline Balance row displays the p-value from an F-test evaluating whether all pre-treatment indicators are simultaneously equal to zero. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

Table 3.10: Summary Statistics by Quintile (Full sample)

		One Employee		1^{st} Quintile		2^{nd} Quintile		3^{rd} Quintile		4^{th} Quintile		5^{th} Quintile
Number of Workers	112113	1.000 (0.000)	17967	2.417 (0.004)	5576	4.000 (0.000)	10385	5.836 (0.008)	11159	11.041 (0.024)	10906	100.319 (3.275)
Gross income declared in 2018	48736	56135.488 (3406.158)	12483	1.37e+05 (15199.471)	4069	1.94e+05 (27108.621)	8080	2.63e+05 (27908.371)	9299	5.18e+05 (25680.115)	9831	6.20e+06 (3.76e+05)
Taxable income declared in 2018	48736	-1815.322 (617.037)	12483	902.351 (1284.135)	4069	3059.471 (3654.404)	8080	5302.446 (1135.026)	9299	5777.748 (3585.382)	9831	3.01e+05 (50463.856)
Declared Exempt Level of Taxes in 2018	48736	0.810 (0.002)	12483	0.537 (0.004)	4069	0.483 (0.008)	8080	0.448 (0.006)	9299	0.398 (0.005)	9831	0.327 (0.005)
Taxes paid in 2018	112113	451.280 (52.816)	17967	1577.032 (209.524)	5576	2398.453 (601.784)	10385	2888.376 (215.448)	11159	5562.824 (307.758)	10906	96080.650 (11414.166)
Sector Unclassified	112113	0.191 (0.001)	17967	0.209 (0.003)	5576	0.191 (0.005)	10385	0.191 (0.004)	11159	0.167 (0.004)	10906	0.165 (0.004)
Primary Sector	112113	0.114 (0.001)	17967	0.125 (0.002)	5576	0.113 (0.004)	10385	0.097 (0.003)	11159	0.097 (0.003)	10906	0.090 (0.003)
Secondary Sector	112113	0.107 (0.001)	17967	0.112 (0.002)	5576	0.122 (0.004)	10385	0.126 (0.003)	11159	0.141 (0.003)	10906	0.208 (0.004)
Tertiary Sector	112113	0.587 (0.001)	17967	0.554 (0.004)	5576	0.573 (0.007)	10385	0.585 (0.005)	11159	0.594 (0.005)	10906	0.538 (0.005)

Note: Table displays the average characteristics for firms in different quintiles of firm size. The number of observations appears to the right of the mean and standard errors appear below the mean in parentheses. This table includes all of the firms in our sample of data with a recorded number of firms. Sample Sizes differ across quintiles because of ties.

Table 3.11: Summary Statistics by Quintile (Experimental Sample)

	One Employee		1^{st} Quintile		2^{nd} Quintile		3^{rd} Quintile		4^{th} Quintile		5^{th} Quintile	
Variable	N	Mean/SE	N	Mean/SE	N	Mean/SE	N	Mean/SE	N	Mean/SE	N	Mean/SE
Number of Workers	36,619	1.000 (0.000)	5,859	2.42 (0.006)	1,823	4.00 (0.00)	3,637	5.85 (0.01)	4,06	11.12 (0.039)	4,306	95.48 (4.6)
Gross income declared in 2018	14,784	76,930.05 (8702.36)	4,202	168,915.90 (15,852.97)	1,359	203,042.40 (20,548.17)	2909	383,531.40 (71,113.23)	3548	753,604.10 (51,490.469)	4,075	7,809,413 (7.23e+05)
Taxable income declared in 2018	14,784	-1,709.432 (1,219.136)	4,202	4,401.614 (3,624.20)	1,359	-2,167.66 (6,143.682)	2,909	10,514.12 (2,559.534)	3548	20,828.836 (3,162.558)	4,075	446,581.5 (78,161.546)
Declared Exempt Level of Taxes in 2018	14,784	0.776 (0.003)	4,202	0.485 (0.008)	1,359	0.444 (0.013)	2,909	0.390 (0.009)	3,548	0.306 (0.008)	4,075	0.213 (0.006)
Taxes paid in 2018	36,619	583.530 (61.537)	5,859	2,652.632 (629.025)	1,823	2,306.815 (348.789)	3,637	4,127.842 (519.664)	4,066	8,242.694 (616.656)	4,306	1.40e+05 (18,443.150)

Note: Table displays the average characteristics for firms in different quintiles of firm size. The number of observations appears to the right of the mean and standard errors appear below the mean in parentheses. This table includes all of the firms in our sample of data with a recorded number of firms who were sent messages. Sample Sizes differ across quintiles because the quintile is calculated using the full sample.

Table 3.12: Treatment Effect Heterogeneity by Firm Size

	(1) Probability of Increasing Taxes Paid	(2) Change in Taxes Paid
Commission × Log Employees	0.00323	-2127.1
	(0.00434)	(4439.1)
Public with Commission × Log Employees	0.0942***	4849.3
	(0.00509)	(3620.3)
Public with Commission × Log Employees	0.109***	5372.1
	(0.00512)	(7412.1)
Prison × Log Employees	0.0972***	22636.6*
	(0.00515)	(11850.5)
Prison with Commission × Log Employees	0.136***	31205.7***
	(0.00508)	(10417.7)
Commission	0.00410	879.8
	(0.00396)	(2047.0)
Public	0.0277***	-1317.6
	(0.00439)	(1600.7)
Public with Commission	0.0390***	-1804.1
	(0.00451)	(3152.8)
Prison	0.0248***	-10511.3*
	(0.00435)	(6186.0)
Prison with Commission	0.0576***	-11388.2**
	(0.00466)	(5090.6)
Log Employees	0.0411***	-2807.1
	(0.00301)	(2521.9)
Constant	0.0535***	793.1
	(0.00276)	(1174.8)
R-squared	0.191	0.012
Observations	56310	56310

Note: Table displays coefficient estimates from regression treatment dummies, log employees and interactions on the probability of increasing taxes paid and the change in taxes paid from 2018 to 2019. Robust standard errors in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

REFERENCES

Abadie, Alberto, “Bootstrap tests for distributional treatment effects in instrumental variable models,” *Journal of the American statistical Association*, 2002, *97* (457), 284–292.

Acemoglu, Daron, “Politics and economics in weak and strong states,” *Journal of monetary Economics*, 2005, *52* (7), 1199–1226.

_ , Camilo García-Jimeno, and James A Robinson, “State capacity and economic development: A network approach,” *American Economic Review*, 2015, *105* (8), 2364–2409.

Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth, “Killer incentives: Relative position, performance and risk-taking among German fighter pilots, 1939-45,” *NBER working paper*, 2018, *22992*, 16.

Aigner, Dennis J, “Regression with a binary independent variable subject to errors of observation,” *Journal of Econometrics*, 1973, *1* (1), 49–59.

Akerlof, George and Janet L Yellen, *Gang behavior, law enforcement, and community values*, Canadian Institute for Advanced Research Washington, DC, 1994.

Amazon, “Price gouging has no place in our stores,” 2020.

Ambuehl, Sandro, Muriel Niederle, and Alvin E Roth, “More money, more problems? Can high pay be coercive and repugnant?,” *American Economic Review*, 2015, *105* (5), 357–60.

Anderson, Eric T and Duncan I Simester, “Price stickiness and customer antagonism,” *The quarterly journal of economics*, 2010, *125* (2), 729–765.

Ang, Desmond, “The effects of police violence on inner-city students,” *The Quarterly Journal of Economics*, 2021, *136* (1), 115–168.

Angrist, Joshua D, “The Perils of Peer Effects,” *Labour Economics*, 2014, *30*, 98–108.

Annan-Phan, Sébastien and Bocar A Ba, “Hot Temperatures, Aggression, and Death at the Hands of the Police: Evidence from the US,” *Aggression, and Death at the Hands of the Police: Evidence from the US (July 3, 2020)*, 2020.

Antinyan, Armenak and Zareh Asatryan, “Nudging for tax compliance: A meta-analysis,” *ZEW-Centre for European Economic Research Discussion Paper*, 2019, (19-055).

Ariel, Barak, “Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial,” *Criminology*, 2012, *50* (1), 27–69.

Ba, Bocar, “Going the extra mile: The cost of complaint filing, accountability, and law enforcement outcomes in Chicago,” Technical Report, Working paper 2017.

_ , “Going the extra mile: The cost of complaint filing, accountability, and law enforcement outcomes in Chicago,” Technical Report, Working paper 2018.

Ba, Bocar A and Roman Rivera, “The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago,” Technical Report, Working Paper 2019.

Ba, Bocar, Dean Knox, Jonathan Mummolo, and Roman Rivera, “The Impact of Racial and Ethnic Diversity in Policing,” Technical Report, Working Paper 2020.

Baird, Sarah, J Aislinn Bohren, Craig McIntosh, and Berk Özler, “Optimal design of experiments in the presence of interference,” *Review of Economics and Statistics*, 2018, *100* (5), 844–860.

Balafoutas, Loukas and Nikos Nikiforakis, “Norm enforcement in the city: A natural field experiment,” *European Economic Review*, 2012, *56* (8), 1773–1785.

_ , _ , and Bettina Rockenbach, “Altruistic punishment does not increase with the severity of norm violations in the field,” *Nature communications*, 2016, *7* (1), 1–6.

Banerjee, Abhijit V., “A Simple Model of Herd Behavior,” *The Quarterly Journal of Economics*, 08 1992, *107* (3), 797–817.

Barzel, Yoram, “A theory of rationing by waiting,” *The Journal of Law and Economics*, 1974, *17* (1), 73–95.

Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts, “Can war foster cooperation?,” *Journal of Economic Perspectives*, 2016, *30* (3), 249–74.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen, “Building criminal capital behind bars: Peer effects in juvenile corrections,” *The Quarterly Journal of Economics*, 2009, *124* (1), 105–147.

Beatty, Timothy, Gabriel Lade, and Jay Shimshack, “Hurricanes and Gasoline Price Gouging,” 2020.

Becker, Gary S, “A Theory of the Allocation of Time,” *The economic journal*, 1965, pp. 493–517.

Bénabou, Roland and Jean Tirole, “Incentives and prosocial behavior,” *American Economic Review*, 2006, *96* (5), 1652–1678.

Bergeron, Augustin, Gabriel Tourek, and Jonathan L Weigel, “The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC.”

Bérgolo, Marcelo L, Rodrigo Ceni, Guillermo Cruces, Matias Giaccobasso, and Ricardo Perez-Truglia, “Tax audits as scarecrows: Evidence from a large-scale field experiment,” Technical Report, National Bureau of Economic Research 2017.

Berry, James, Greg Fischer, and Raymond Guiteras, “Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana,” *Journal of Political Economy*, 2020, *128* (4), 1436–1473.

Besley, Timothy and Torsten Persson, "The origins of state capacity: Property rights, taxation, and politics," *American economic review*, 2009, *99* (4), 1218–44.

_ and _ , "Why do developing countries tax so little?," *Journal of Economic Perspectives*, 2014, *28* (4), 99–120.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Black, Dan A, Mark C Berger, and Frank A Scott, "Bounding parameter estimates with nonclassical measurement error," *Journal of the American Statistical Association*, 2000, *95* (451), 739–748.

Blau, Robert and John Kass, "POLICE HIRING LOTTERY LATEST DALEY HEADACHE," Sep 2018.

Bø, Erlend E, Joel Slemrod, and Thor O Thoresen, "Taxes on the internet: Deterrence effects of public disclosure," *American Economic Journal: Economic Policy*, 2015, *7* (1), 36–62.

Bobo, Lawrence D and Victor Thompson, "Unfair by design: The war on drugs, race, and the legitimacy of the criminal justice system," *Social Research: An International Quarterly*, 2006, *73* (2), 445–472.

Brock, William A and Steven N Durlauf, "Discrete Choice with Social Interactions," *The Review of Economic Studies*, 2001, *68* (2), 235–260.

Bronfenbrenner, Martin, "Price control under imperfect competition," *The American Economic Review*, 1947, *37* (1), 107–120.

Brown, Ryan, Verónica Montalva, Duncan Thomas, and Andrea Velásquez, "Impact of Violent Crime on Risk Aversion: Evidence from the Mexican Drug War," *Review of Economics and Statistics*, 2019, *101* (5), 892–904.

Brunson, Rod K. and Jody Miller, "Young Black Men and Urban Policing in the United States," *The British Journal of Criminology*, 10 2005, *46* (4), 613–640.

Bursztyn, Leonardo and Robert Jensen, "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure," *Annual Review of Economics*, 2017, *9*, 131–153.

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky, "The deadweight loss of social recognition," Technical Report, National Bureau of Economic Research 2019.

Cabral, Luis and Lei Xu, "Seller Reputation and Price Gouging: Evidence from the COVID-19 Pandemic," *Mimeo, April*, 2020.

Callen, Michael, Mohammad Isaqzadeh, James D Long, and Charles Sprenger, "Violence and Risk Preference: Experimental Evidence from Afghanistan," *American Economic Review*, 2014, *104* (1), 123–48.

Cameron, Lisa and Manisha Shah, "Risk-Taking Behavior in the Wake of Natural Disasters," *Journal of Human Resources*, 2015, *50* (2), 484–515.

Card, David and Gordon B Dahl, "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *The Quarterly Journal of Economics*, 2011, *126* (1), 103–143.

Carlsmith, Kevin M, John M Darley, and Paul H Robinson, "Why do we punish? Deterrence and just deserts as motives for punishment.," *Journal of personality and social psychology*, 2002, *83* (2), 284.

Carpio, Lucia Del, "Are the neighbors cheating? Evidence from a social norm experiment on property taxes in Peru," *Princeton, NJ, Princeton University Working Paper*, 2013.

Carrell, Scott E and Mark L Hoekstra, "Externalities in the classroom: How children exposed to domestic violence affect everyone's kids," *American Economic Journal: Applied Economics*, 2010, *2* (1), 211–28.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

Cavallo, Alberto, Eduardo Cavallo, and Roberto Rigobon, "Prices and supply disruptions during natural disasters," *Review of Income and Wealth*, 2014, *60*, S449–S471.

CFD, Aug 2014.

Chakraborti, Rik and Gavin Roberts, "Anti-Price Gouging Laws, Shortages, and COVID-19: Big Data Insights from Consumer Searches," *SSRN Working Paper*, 2020.

_ and _ , "Learning to Hoard: the Effects of Preexisting and Surprise Price-Gouging Regulation during the COVID-19 Pandemic," *SSRN Working Paper*, 2020.

Chalfin, Aaron and Justin McCrary, "Are US cities underpoliced? Theory and evidence," *Review of Economics and Statistics*, 2018, *100* (1), 167–186.

Charness, Gary and Matthias Sutter, "Groups make better self-interested decisions," *Journal of Economic Perspectives*, 2012, *26* (3), 157–76.

Clemens, Michael A, "Testing for repugnance in economic transactions: Evidence from guest work in the Gulf," *The Journal of Legal Studies*, 2018, *47* (S1), S5–S44.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg, "Peer effects in the workplace," *American Economic Review*, 2017, *107* (2), 425–56.

CPD, "ENTRY LEVEL POLICE OFFICER FREQUENTLY ASKED QUESTIONS ("FAQS")," 2018.

Cushman, Fiery, "Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment," *Cognition*, 2008, *108* (2), 353–380.

de Chaisemartin, Clément and Luc Behaghel, "Estimating the Effect of Treatments Allocated by Randomized Waiting Lists," *Econometrica*, 2020, *88* (4), 1453–1477.

de Quidt, Jonathan, Lise Vesterlund, and Alistair Wilson, "Experimenter demand effects," in Aljaz Ule and Arthur Schram, eds., *Handbook of Research Methods and Applications in Experimental Economics*, Cheltenham, UK: Edward Elgar Publishing, 2019, chapter 20, pp. 384–400.

Denant-Boemont, Laurent, Enrico Diecidue, and Olivier L'haridon, "Patience and time consistency in collective decisions," *Experimental economics*, 2017, *20* (1), 181–208.

DeScioli, Peter, John Christner, and Robert Kurzban, "The omission strategy," *Psychological Science*, 2011, *22* (4), 442–446.

Diekmann, Andreas, "Volunteer's dilemma," *Journal of conflict resolution*, 1985, *29* (4), 605–610.

Dincecco, Mark and Gabriel Katz, "State capacity and long-run economic performance," *The Economic Journal*, 2016, *126* (590), 189–218.

DiPasquale, Denise and Edward L Glaeser, "The Los Angeles riot and the economics of urban unrest," *Journal of Urban Economics*, 1998, *43* (1), 52–78.

Dirección General de Impuestos Internos, "Memorias 2016-2020," 2020.

Draca, Mirko, Stephen Machin, and Robert Witt, "Panic on the streets of london: Police, crime, and the july 2005 terror attacks," *American Economic Review*, 2011, *101* (5), 2157–81.

Du, Shuili and Edward T Vieira, "Striving for legitimacy through corporate social responsibility: Insights from oil companies," *Journal of Business Ethics*, 2012, *110* (4), 413–427.

Dwenger, Nadja and Lukas Treber, "Shaming for tax enforcement: Evidence from a new policy," 2018.

Dworczak, Piotr, Scott Duke Kominers, and Mohammad Akbarpour, "Redistribution through markets," *Becker Friedman Institute for Research in Economics Working Paper*, 2019, (2018-16).

Edwards, Frank, Hedwig Lee, and Michael Esposito, "Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex," *Proceedings of the National Academy of Sciences*, 2019, *116* (34), 16793–16798.

Elías, Julio J., Nicola Lacetera, and Mario Macis, “Paying for Kidneys? A Randomized Survey and Choice Experiment,” *American Economic Review*, August 2019, *109* (8), 2855–88.

Eren, Ozkan and Naci Mocan, “Emotional judges and unlucky juveniles,” *American Economic Journal: Applied Economics*, 2018, *10* (3), 171–205.

Evans, William N and Emily G Owens, “COPS and Crime,” *Journal of Public Economics*, 2007, *91* (1-2), 181–201.

Fehr, Ernst and Simon Gächter, “Altruistic punishment in humans,” *Nature*, 2002, *415* (6868), 137–140.

_ and Urs Fischbacher, “Third-party punishment and social norms,” *Evolution and human behavior*, 2004, *25* (2), 63–87.

_ , _ , and Simon Gächter, “Strong reciprocity, human cooperation, and the enforcement of social norms,” *Human nature*, 2002, *13* (1), 1–25.

Fryer, Roland, “A Response to Steven Durlauf and James Heckman,” 2020.

G., Jr. Fryer Roland, “Reconciling Results on Racial Differences in Police Shootings,” *AEA Papers and Proceedings*, May 2018, *108*, 228–33.

Garlick, Robert, “Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment,” *American Economic Journal: Applied Economics*, 2018, *10* (3), 345–369.

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

Ginther, Matthew R, Richard J Bonnie, Morris B Hoffman, Francis X Shen, Kenneth W Simons, Owen D Jones, and René Marois, “Parsing the behavioral and brain mechanisms of third-party punishment,” *Journal of Neuroscience*, 2016, *36* (36), 9420–9434.

Giosa, Penelope, “Exploitative Pricing in the Time of Coronavirus—The Response of EU Competition Law and the Prospect of Price Regulation,” *Journal of European Competition Law & Practice*, 2020, *11* (9), 499–508.

Gobierno de la República Dominicana: Equipo Interinstitucional, “Estimación del Incumplimiento Tributario en la República Dominicana,” 2018.

Gonzalez, Richard and George Wu, “On the Shape of the Probability Weighting Function,” *Cognitive Psychology*, 1999, *38* (1), 129 – 166.

Guryan, Jonathan, Kory Kroft, and Matthew J Notowidigdo, “Peer effects in the workplace: Evidence from random groupings in professional golf tournaments,” *American Economic Journal: Applied Economics*, 2009, *1* (4), 34–68.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman, "Tax Evasion at the Top of the Income Distribution: Theory and Evidence," 2021.

Hahn, Jinyong and Whitney Newey, "Jackknife and analytical bias reduction for nonlinear panel models," *Econometrica*, 2004, *72* (4), 1295–1319.

Hallsworth, Michael, John A List, Robert D Metcalfe, and Ivo Vlaev, "The making of homo honoratus: From omission to commission," Technical Report, National Bureau of Economic Research 2015.

Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe, "Do risk preferences change? evidence from the great east japan earthquake," *American Economic Journal: Applied Economics*, 2018, *10* (2), 298–330.

Hanlon, Michelle and Joel Slemrod, "What does tax aggressiveness signal? Evidence from stock price reactions to news about tax shelter involvement," *Journal of Public Economics*, 2009, *93* (1-2), 126–141.

Harrison, Glenn, "Theory and Misbehavior of First-Price Auctions: Reply," *American Economic Review*, 1992, *82* (5), 1426–43.

Harrison, Glenn W and John A List, "Field experiments," *Journal of Economic literature*, 2004, *42* (4), 1009–1055.

Hasegawa, Makoto, Jeffrey L Hoopes, Ryo Ishida, and Joel B Slemrod, "The effect of public disclosure on reported taxable income: Evidence from individuals and corporations in Japan," *Available at SSRN 1653948*, 2012.

He, Hongwei and Lloyd Harris, "The impact of Covid-19 pandemic on corporate social responsibility and marketing philosophy," *Journal of Business Research*, 2020, *116*, 176–182.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in "Proceedings of the IEEE conference on computer vision and pattern recognition" 2016, pp. 770–778.

Heckman, James J and Steven N Durlauf, "Comment on" An Empirical Analysis of Racial Differences in Police Use of Force" by Roland G. Fryer Jr.," 2020.

Hedblom, Daniel, Brent R Hickman, and John A List, "Toward an Understanding of Corporate Social Responsibility: Theory and Field Experimental Evidence," Technical Report, National Bureau of Economic Research 2019.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich et al., "Costly punishment across human societies," *Science*, 2006, *312* (5781), 1767–1770.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter, "Antisocial punishment across societies," *Science*, 2008, *319* (5868), 1362–1367.

Herrnstadt, Evan, Anthony Heyes, Erich Muehlegger, and Soodeh Saberian, "Air pollution as a cause of violent crime: Evidence from Los Angeles and Chicago," *Manuscript in preparation*, 2016.

Hirano, Keisuke and Jinyong Hahn, "Design of randomized experiments to measure social interaction effects," *Economics Letters*, 2010, *106* (1), 51–53.

Hjort, Jonas, "Ethnic divisions and production in firms," *The Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.

Holshue, Michelle L., Chas DeBolt, Scott Lindquist, Kathy H. Lofy, John Wiesman, Hollianne Bruce, Christopher Spitters, Keith Ericson, Sara Wilkerson, Ahmet Tural, George Diaz, Amanda Cohn, LeAnne Fox, Anita Patel, Susan I. Gerber, Lindsay Kim, Suxiang Tong, Xiaoyan Lu, Steve Lindstrom, Mark A. Pallansch, William C. Weldon, Holly M. Biggs, Timothy M. Uyeki, and Satish K. Pillai, "First Case of 2019 Novel Coronavirus in the United States," *New England Journal of Medicine*, 2020, *382* (10), 929–936. PMID: 32004427.

Hoopes, Jeffrey L, Leslie Robinson, and Joel Slemrod, "Public tax-return disclosure," *Journal of Accounting and Economics*, 2018, *66* (1), 142–162.

Hotz, V Joseph, Guido W Imbens, and Julie H Mortimer, "Predicting the efficacy of future training programs using past experiences at other locations," *Journal of Econometrics*, 2005, *125* (1-2), 241–270.

Imas, Alex, Michael A Kuhn, and Vera Mironova, "Exposure to Violence Predicts Impulsivity in Time Preference: Evidence from The Democratic Republic of Congo," Technical Report, Working Paper 2018.
Joe Biden

Joe Biden*, Oct 2020.*

Kahneman, Daniel, Jack L Knetsch, and Richard Thaler, "Fairness as a constraint on profit seeking: Entitlements in the market," The American economic review, 1986, pp. 728–741.

Kaldor, Nicholas, "Taxation for economic development," The Journal of Modern African Studies, 1963, 1 (1), 7–23.

Kamerow, Douglas, "Covid-19: the crisis of personal protective equipment in the US," BMJ, 2020, 369.

Kettle, Stewart, Marco Hernandez, Simon Ruda, and Michael A Sanderson, "Behavioral Interventions in Tax Compliance: Evidence from Guatemala," World Bank Policy Research Working Paper, 2016, (7690).

Kirk, David S and Andrew V Papachristos, "Cultural mechanisms and the persistence of neighborhood violence," American journal of sociology, 2011, 116 (4), 1190–1233.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "Why can modern governments tax so much? An agency model of firms as fiscal intermediaries," Economica, 2016, 83 (330), 219–246.

_ , Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez, "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark," Econometrica, 2011, 79 (3), 651–692.

Kőszegi, Botond, "Emotional agency," The Quarterly Journal of Economics, 2006, 121 (1), 121–155.

Kumler, Todd, Eric Verhoogen, and Judith A Frías, "Enlisting employees in improving payroll-tax compliance: Evidence from Mexico," Technical Report, National Bureau of Economic Research 2013.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva, "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments," American Economic Review, April 2015, 105 (4), 1478–1508.

Lartey, Jamiles, "By the numbers: US police kill more in days than other countries do in years," The Guardian, 2015, 9.

Latinobarómetro Corporation, "Latinobarómetro Databank," 2016.

Lee, David S and Justin McCrary, "The Deterrence Effect of Prison: Dynamic Theory and Evidence," in "Regression Discontinuity Designs (Advances in Econometrics)," Vol. 38, Emerald Publishing Limited, 2017.

Legewie, Joscha and Jeffrey Fagan, "Aggressive policing and the educational performance of minority youth," American Sociological Review, 2019, 84 (2), 220–247.

Levitt, Steven D, "Using electoral cycles in police hiring to estimate the effects of police on crime: Reply," American Economic Review, 2002, 92 (4), 1244–1250.

List, John A, "Does market experience eliminate market anomalies?," The Quarterly Journal of Economics, 2003, 118 (1), 41–71.

_ , "Does market experience eliminate market anomalies? The case of exogenous market experience," American Economic Review, 2011, 101 (3), 313–17.

_ , "Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial," Technical Report, National Bureau of Economic Research 2020.

_ and Fatemeh Momeni, "When corporate social responsibility backfires: Theory and evidence from a natural field experiment," Technical Report, National Bureau of Economic Research 2017.

List, John, Daniel Millimet et al., "Bounding the impact of market experience on rationality: Evidence from a field experiment with imperfect compliance," IDEAS Working Paper, 2005, (505).

Loewenstein, George, "Out of control: Visceral influences on behavior," Organizational Behavior and Human Decision Processes, 1996, 65 (3), 272–292.

_ and Ted O'Donoghue, "The heat of the moment: Modeling interactions between affect and deliberation," Unpublished manuscript, 2007, pp. 1–69.

Lum, Cynthia and Daniel S Nagin, "Reinventing american policing," Crime and Justice, 2017, 46 (1), 339–393.

Luttmer, Erzo FP and Monica Singhal, "Tax morale," Journal of Economic Perspectives, 2014, 28 (4), 149–68.

MacIntyre, C Raina, Holly Seale, Tham Chi Dung, Nguyen Tran Hien, Phan Thi Nga, Abrar Ahmad Chughtai, Bayzidur Rahman, Dominic E Dwyer, and Quanyi Wang, "A cluster randomised trial of cloth masks compared with medical masks in healthcare workers," BMJ open, 2015, 5 (4), e006577.

Manski, Charles F, "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon," Economics of Education review, 1989, 8 (4), 305–312.

_ , "Identification of endogenous social effects: The reflection problem," The Review of Economic Studies, 1993, 60 (3), 531–542.

_ and Daniel S Nagin, "Assessing benefits, costs, and disparate racial impacts of confrontational proactive policing," Proceedings of the National Academy of Sciences, 2017, 114 (35), 9308–9313.

Marcelo, Darwin, Aditi Raina, and Stuti Rawat, "Private Sector Participation in Disaster Recovery and Mitigation [Disaster Recovery Guidance Series by the Global Facility for Disaster Reduction and Recovery]," 01 2020.

Marmaros, David and Bruce Sacerdote, "How Do Friendships Form?," The Quarterly Journal of Economics, 2006, 121 (1), 79–119. Publisher: Oxford University Press.

Mas, Alexandre, "Pay, reference points, and police performance," The Quarterly Journal of Economics, 2006, 121 (3), 783–821.

_ and Enrico Moretti, "Peers at work," American Economic Review, 2009, 99 (1), 112–45.

Mascagni, Giulia, Christopher Nell, and Nara Monkam, "One size does not fit all: a field experiment on the drivers of tax compliance and delivery methods in Rwanda," 2017.

McDonnell, Mary-Hunter and Brayden King, "Keeping up appearances: Reputational threat and impression management after social movement boycotts," Administrative Science Quarterly, 2013, 58 (3), 387–419.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook, "Birds of a feather: Homophily in social networks," Annual review of sociology, 2001, 27 (1), 415–444.

Morales, Miguel, "The Economic Value of Crime Control: Evidence from a Large Investment on Police Infrastructure in Colombia," Technical Report, Working Paper 2020.

Moss, Aaron J, Cheskie Rosenzweig, Jonathan Robinson, and Leib Litman, "Is it ethical to use Mechanical Turk for behavioral research? Relevant data from a representative survey of MTurk participants and wages," 2020.

Moya, Andrés, "Violence, psychological trauma, and risk attitudes: Evidence from victims of violence in Colombia," American Journal of Preventive Medicine, 2018, 54, 503–509.

Munyo, Ignacio and Martín Antonio Rossi, "The effects of real exchange rate fluctuations on the gender wage gap and domestic violence in Uruguay," Technical Report, IDB Working Paper Series 2015.

Murphy, Francis X, "Does Increased Exposure to Peers with Adverse Characteristics Reduce Workplace Performance? Evidence from a Natural Experiment in the US Army," Journal of Labor Economics, 2019, 37 (2), 435–466.

NBC, "Thousands Take Entry-Level Police Exam," Dec 2013.

Neve, Jan-Emmanuel De, Clement Imbert, Johannes Spinnewijn, Teodora Tsankova, and Maarten Luts, "How to improve tax compliance? Evidence from population-wide experiments in Belgium," Evidence from Population-Wide Experiments in Belgium (May 05, 2019). Saïd Business School WP, 2019, 7.

Neyman, Jerzy and Elizabeth L Scott, "Consistent estimates based on partially consistent observations," Econometrica: Journal of the Econometric Society, 1948, pp. 1–32.

of Justice Press Release 15057, Department, "Justice Department Announces Findings of Investigation into Chicago Police Department," Justice News, Jan 2017.

OIG, Police Accountability Task Force, Apr 2016.

Osofsky, Joy D, "The effects of exposure to violence on young children (1995).," in "Carnegie Corporation of New York Task Force on the Needs of Young Children; An earlier version of this article was presented as a position paper for the aforementioned corporation." American Psychological Association 1997.

Owens, Emily, David Weisburd, Karen L Amendola, and Geoffrey P Alpert, "Can you build a better cop? Experimental evidence on supervision, training, and policing in the community," Criminology & Public Policy, 2018, 17 (1), 41–87.

Pedersen, Eric J, William HB McAuliffe, and Michael E McCullough, "The unresponsive avenger: More evidence that disinterested third parties do not punish altruistically.," Journal of Experimental Psychology: General, 2018, 147 (4), 514.

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

Philipson, Tomas, "Economic epidemiology and infectious diseases," Handbook of health economics, 2000, 1, 1761–1799.

Pigou, A. C., The economics of welfare / by A. C. Pigou, Macmillan London, 1920.

Pomeranz, Dina, "No taxation without information: Deterrence and self-enforcement in the value added tax," American Economic Review, 2015, 105 (8), 2539–69.

Pritchard, Paige, "Do You Have What It Takes To Join the Chicago Police Department?," Aug 2013.

Rauch, James E and Peter B Evans, "Bureaucratic structure and bureaucratic performance in less developed countries," Journal of public economics, 2000, 75 (1), 49–71.

Rey, Jason Del, "Amazon is banning the sale of N95 and surgical masks to the general public," Vox.com, Apr 2020.

Rick, Scott and George Loewenstein, "The role of emotion in economic behavior," Handbook of emotions, 2008, 3, 138–158.

Rohlfs, Chris, "Does combat exposure make you a more violent or criminal person? Evidence from the Vietnam draft," Journal of Human Resources, 2010, 45 (2), 271–300.

Rotemberg, Julio J, "Customer anger at price increases, changes in the frequency of price adjustment and monetary policy," Journal of monetary economics, 2005, 52 (4), 829–852.

_ , "Behavioral aspects of price setting, and their policy implications," Technical Report, National Bureau of Economic Research 2008.

_ , "Fair pricing," Journal of the European Economic Association, 2011, 9 (5), 952–981.

Roth, Alvin E, "Repugnance as a Constraint on Markets," Journal of Economic perspectives, 2007, 21 (3), 37–58.

_ , Who gets what—and why: The new economics of matchmaking and market design, Houghton Mifflin Harcourt, 2015.

Sacerdote, Bruce, "Peer effects with random assignment: Results for Dartmouth roommates," The Quarterly journal of economics, 2001, 116 (2), 681–704.

Sandler, Danielle H and Ryan Sandler, "Multiple event studies in public finance and labor economics: A simulation study with applications," Journal of Economic and Social Measurement, 2014, 39 (1-2), 31–57.

Selyukh, Alina, 'Stop Price Gouging,' 33 Attorneys General Tell Amazon, Walmart, Others, Mar 2020.

Servaes, Henri and Ane Tamayo, "The impact of corporate social responsibility on firm value: The role of customer awareness," Management Science, 2013, 59 (5), 1045–1061.

Shue, Kelly, "Executive networks and firm policies: Evidence from the random assignment of MBA peers," The Review of Financial Studies, 2013, 26 (6), 1401–1442.

Slemrod, Joel, "Tax compliance and enforcement: New research and its policy implications," 2016.

_ , "Tax compliance and enforcement," Journal of Economic Literature, 2019, 57 (4), 904–54.

_ , Marsha Blumenthal, and Charles Christian, "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota," Journal of Public Economics, 2001, 79 (3), 455–483.

Smith, Adam, "An inquiry into the wealth of nations," Strahan and Cadell, London, 1776, pp. 1–11.

Spranca, Mark, Elisa Minsk, and Jonathan Baron, "Omission and commission in judgment and choice," Journal of Experimental Social Psychology, 1991, 27 (1), 76–105.

Ta, Vivian P, Brian Lande, and Joel Suss, "Emotional Reactivity and Police Expertise in Use-of-Force Decision-Making," Journal of Police and Criminal Psychology, 2021, pp. 1–10.

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen, "Risk and time preferences: Linking experimental and household survey data from Vietnam," American Economic Review, 2010, 100 (1), 557–71.

Tankersley, Jim, "Live Updates: Biden to Toughen Tax Enforcement to Help Pay for His Economic Agenda," Apr 2021.

Tella, Rafael Di and Ernesto Schargrodsky, "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack," American Economic Review, 2004, 94 (1), 115–133.

Thaler, Richard H and Cass R Sunstein, Nudge: Improving decisions about health, wealth, and happiness, Penguin, 2009.

Thomson-DeVeaux, Amelia, Laura Bronner, and Damini Sharma, "Cities Spend Millions On Police Misconduct Every Year. Here's Why It's So Difficult to Hold Departments Accountable.," Feb 2021.

Tiesman, Hope, Melody, Srinivas Gwilliam, Jeff Konda, Rojek, and Suzanne Marsh, "Nonfatal Injuries to Law Enforcement Officers: A Rise in Assaults," Review of Economics and Statistics, 2019, 101 (5), 892–904.

Tobin, James, "On limiting the domain of inequality," The Journal of Law and Economics, 1970, 13 (2), 263–277.

Tong, Lester CP, J Ye Karen, Kentaro Asai, Seda Ertac, John A List, Howard C Nusbaum, and Ali Hortaçsu, "Trading experience modulates anterior insula to reduce the endowment effect," Proceedings of the National Academy of Sciences, 2016, 113 (33), 9238–9243.

Tyler, Tom R, "Enhancing police legitimacy," The Annals of the American Academy of Political and Social Science, 2004, 593 (1), 84–99.

Uber, "Uber y la CDMX acuerdan tarifas durante contingencias," `https://www.uber.com/es-MX/blog` *2016.*

Weitzer, Ronald and Steven A Tuch, "Race and perceptions of police misconduct," Social Problems, 2004, 51 (3), 305–325.

Weitzman, Martin L, "Price distortion and shortage deformation, or what happened to the soap?," The American Economic Review, 1991, pp. 401–414.

Wooldridge, Jeffrey M, "Distribution-free estimation of some nonlinear panel data models," Journal of Econometrics, 1999, 90 (1), 77–97.

Zimring, Franklin E, "What Drives Variation in Killings by Urban Police in the United States: Two Empirical Puzzles," The Cambridge Handbook of Policing in the United States, 2019, p. 296.