Police are less likely to respond to requests for help from minorities: Field experiment evidence of police discrimination[*†]

latest draft

Garrett Olson Stanford
Department of Economics, 736 PLC
University of Oregon
Eugene, OR 97403-1285
gos@uoregon.edu

January 2023

Abstract

Numerous high-profile incidents have led to accusations that police departments struggle with equity and accountability. I use an experiment to test both. In the form of a correspondence study, I send emails to more than 2,000 U.S. police departments requesting information about how to lodge a complaint against an officer. Manipulating the names of the email sender, I compare department response rates across race/ethnicity (Black, Hispanic and White) and gender (female and male). I find that departments are less likely to respond to emails signed with Black and Hispanic names. Differences in response rates become more pronounced when I interact gender with race/ethnicity. These differences exacerbate a low overall response rate of 67.4 percent. I find little evidence that department size or the local population are correlated with response rates. Results from this experiment support the accusations that policing suffers from issues of bias and transparency.

Keywords: Policing, Police Discrimination, Police Accountability, Racial Justice, Criminal Justice, Field Experiment
JEL Classification: C93, J15, J16, K40

[*]I thank Ed Rubin, Jon Davis and Trudy Ann Cameron for invaluable input and guidance. I also thank my peers at the University of Oregon Microeconomics Group for helpful feedback. Finally, I am especially grateful to Joe Mitchell-Nelson for the inspiration and encouragement to undertake this project.

[†]This study was preregistered with the AEA RCT registry. The pre-analysis plan can be found here.

1 Overview

Research suggests that police likely generate substantial benefits to society. Benefits can accrue from the direct effect of reducing crime (e.g., Chalfin et al., 2021; Mello, 2019; Weisburst, 2019b; Weisburd, 2021). Social benefits also can result from less direct interventions, like a reduction in traffic fatalities (e.g., DeAngelo and Hansen, 2014). Chalfin and McCrary (2018) argue that the benefits of policing are so considerable that despite their substantial budget, many departments remain underfunded. However, there is a longstanding debate concerning the impacts that police practices may have on social welfare. Of primary concern is the presence of bias in police behavior—particularly racially motivated bias. In 2020, protests over racially biased policing broke out across the nation after officers killed George Floyd in the Minneapolis police department. Despite the constant presence of police reform in the national dialogue and alarmingly frequent anecdotes reported in the media, few studies exist that causally document bias in the police force (Smith et al., 2017).

Biased policing is grounded on the premise that police interact with citizens at a different frequency or in a different manner depending on the sociodemographic characteristics of citizens. Police dictate the frequency of interaction in their decisions of where to police (e.g., patrol routes) and whom they police (e.g., traffic stops). Conditional on an interaction taking place, the manner in which police conduct themselves can vary in terms of treatment during the interaction (e.g., use of force), the resulting outcome (e.g., citation or arrest), or police accountability following an interaction (e.g., consequences for police misconduct).

A growing body of research highlights the disproportionate burden policing can place on people of color in these various avenues. For example, Chen et al. (2021) find that neighborhoods with larger Black populations experience a considerably higher police presence. Similarly, there is evidence that people of color are more likely to be stopped by police (e.g., Bulman, 2019; Gelman et al., 2007; Pierson et al., 2020). There is also evidence that the result of a police-citizen interaction depends on the citizen's race. Numerous studies find that people of color are more likely to experience police use of force (e.g., Edwards et al., 2019; Fryer, 2020; Nix et al., 2017; Ross, 2015). People of color are also more likely to be targeted for traffic citations and asset forfeitures (e.g., Goncalves and Mello, 2021; Makowsky et al., 2019; Sances and You, 2017; West, 2018). Research remains limited on biased police accountability. Stroube (2021) documents that in Chicago formal complaints made by Black residents were less likely to be sustained than formal complaints made by White residents. However, most studies that address biased policing, while descriptively informative, cannot attribute causality.

Establishing causality in the context of biased policing is difficult. First, differences in the frequency of interaction do not necessarily reflect biased policing. There is the possibility of a selection issue. Consider the scenario where sociodemographic groups participate in criminal activity at different frequencies. In this case, unbiased policing could still result in heterogeneous rates of police-citizen interactions across sociodemographic groups (Fridell, 2017). Second, measuring biased policing by comparing outcomes for citizens during an interaction with police does not permit causal inference. Suppose the motivation for police initiating an interaction with a citizen is biased, and outcomes for all police-citizen interactions are similar. In that case, this form of analysis will obscure the presence of biased policing (e.g., Knox et al., 2020; Ross et al., 2018). Causal identification of bias in police accountability is especially challenging. Leveraging observational data necessitates substantial assumptions. Consider comparing differences in sanctions for officers between White and Black complainants. One must assume that the reason for the interaction, the conduct during the interaction, and the actions taken by the complainant after the interaction are approximately identical. Furthermore, to compare formal complaint outcomes, they need to be filed, which could be a substantial obstacle (Ba, 2016). Thus, researchers remain divided on the existence and extent of biased policing (Smith et al., 2017; Fridell, 2017).[1] Furthermore, while some researchers employ research designs that permit causal inferences, they make strong assumptions and the vast majority of these studies rely on self-reported police data. Such reliance on police-reported data can lead to inconclusive or incorrect conclusions if police departments strategically or unintentionally misreport (e.g., Luh, 2019).

In this paper, I causally estimate the effect of bias on police transparency. To identify this effect, I conducted a field experiment on a sample of 2,134 U.S. police departments. I created six fictitious citizen identities with three races (Black, Hispanic and White) and two genders (male and female). To create an identity, I chose first and last names that are strongly associated with a specific race/gender and commonly used and created individual email accounts for the identities. I then used these identities to email each

[1] While limited, research has made efforts to address these challenges and find evidence of biased policing (e.g., Antonovics and Knight, 2009; Gelman et al., 2007; Ross, 2015; West, 2018).

department a request for help making a complaint about an officer in the department.[2] I sent each police department an identical email, with two exceptions. First, I varied the email sender's name to signal race and gender. Second, I varied the signoff of the email between an amicable and curt tone. Police departments responded to 67.4 percent of the requests. Response rates for emails from Black or Hispanic identities were both 10 percentage points (pp) lower than the response rates for emails from White identities—differences that are both significant at the 1% level. Emails from White male identities received the highest response rate, at 75.8%, marginally higher than the response rate for emails from White female identities. Response rates signed with Black and Hispanic male identities were 13.9 and 15 pp lower than White male identities and were marginally lower than Black and Hispanic female identities response rates. The signoff of the email did not affect response rates in aggregate or when interacted with identities.

I use this particular experimental design for several reasons. The challenge of causally identifying discrimination is not unique to the context of law enforcement. Over the last decade, correspondence studies, a type of randomized controlled trial (RCT), have become an increasingly popular tool for researchers studying the presence of discrimination (Bertrand and Duflo, 2017).[3] Emulating the seminal work of Bertrand and Mullainathan (2004), researchers use correspondence studies to identify a variety of types of discrimination (e.g., gender, age, or race) in various contexts (e.g., housing, medical services). To date, most correspondence studies focus on Black versus White discrimination, primarily in the context of hiring practices. There have only been a few audit and correspondence studies that focus on discrimination in the provision of public services in the United States (Butler and Broockman, 2011; Einstein and Glick, 2017; Oberfield and Incantalupo, 2021; White et al., 2015). Considering that marginalized groups, on average, are more likely to depend on public services, discriminatory practices are of utmost concern for social planners.

To the best of my knowledge, the only prior correspondence study that concerns law enforcement agencies is Giulietti et al. (2019). These authors conducted a correspondence study with a wide range of public institutions. Included in their list of public institutions are sheriff's offices. In their study, the authors email

[2] In this paper, I will use "race" as a catchall term for both race (Black and White) and ethnicity (Hispanic). I am sensitive to the distinction between race and ethnicity, and I chose to collapse the distinction in this study for simplicity's sake.

[3] In a correspondence study, individuals (often fictitious) who are identical for all observable characteristics other than the characteristic of interest apply for a job, service, or good. The researcher then examines whether the experimentally varied characteristic of interest affects the outcome of the application (Bertrand and Duflo, 2017). The present study uses email instead of the traditional approach of "snail mail," and—as explained below—to request assistance from police departments instead of applying for jobs or making purchases.

the various public institutions with benign requests for information. The authors vary the identity of the requesters, using two distinctively black male names and two distinctively white male names. The authors find that these public institutions (ranging from public libraries to sheriff's offices) are less likely to respond to emails from individuals with distinctively black names. Furthermore, this effect is most pronounced among the various institutions for sheriff's offices.

By using a correspondence study, I overcome two of the main challenges of studying discrimination in the context of law enforcement: (1) finding causal estimates (as opposed to associations) and (2) avoiding potentially compromised self-reporting of data collected or provided by law enforcement agencies. Estimates of properly randomized correspondence studies can be reasonably assumed to be causal. As mentioned above, the primary obstacles to causal inference in biased policing arise from selecting into criminal activities and police discretion over with whom they interact. I avoid these challenges by creating a citizen-initiated police interaction not predicated on a crime taking place and by designating my outcome of interest as the police department's decision to interact.

Avoiding the use of data provided by police departments has significant advantages. First, departments, intentionally or inadvertently, can have ongoing difficulties reporting accurate data (e.g., Luh, 2019)). Second, police data can be a product of subjective reporting by individual police officers and department-specific classification standards.[4] Even when police officers honestly record officer conduct, decisions made in the heat of the moment during a citizen-officer interaction could influence how events are recorded. Finally, departments may be unwilling to disclose "sensitive information."[5]

A correspondence study also allows me to use a nationally representative sample of police departments. I use a sample of over 2,000 police departments representing all states except Hawaii.[6] As a result, the measures of biased policing and accountability from this study represent policing in the United States rather than a specific state, county, or city. Consequently, inferences made in this study are more likely to reflect systemic behavior rather than specific department cultures.

There are two motivations for requesting a complaint form as the experiment's intervention. To start

[4]For instance, PolicingProject.org describes the discrepancies in officer-initiated stop report requirements across states. RevealNews.org finds that the Washington D.C. police department has a comparatively liberal definition of "resisting arrest".)

[5]Weisburst (2019a) does not find evidence of racially biased policing in Dallas. However, Weisburst hypothesizes that the department's willingness to disclose its data to researchers might stem from the fact that the Dallas police do not appear to have a problem with biased policing.

[6]Hawaii's exclusion was a result of random selection. Hawaii only has four distinct police departments.

with, I need a plausible reason for interaction to conduct a correspondence study on police departments. As emergency responders, ethical *and* natural motives for citizen-initiated contact with the police are limited. Requesting a complaint form is particularly well-suited for a correspondence study. The average citizen's familiarity likely does not know how to lodge a formal complaint, and it is believable that a citizen would need guidance in the process. Additionally, it follows that a citizen who believes the police have wronged them would be unwilling or reluctant to talk on the phone or appear at the station in-person (e.g., Ba, 2016).

The primary motivation for requesting a complaint form is to examine police accountability. Research remains limited on police accountability despite its centrality to conversations surrounding equitable policing. By asking departments for help in making a complaint about an officer in their department, I explicitly test the willingness of departments to hold their officers accountable. I provide two measures of police accountability. First, the overall response rate from this experiment is a descriptive measure reflecting a citizen's likelihood of receiving assistance in the complaint process. Second, comparing response rates between the different races and genders of the requesters, I measure the presence of bias in the context of police accountability. Because citizen complaints are one of the only tools available for citizens to address police misconduct, it is crucial to understand the obstacles to this—especially racial or gender discrimination.

In addition to establishing causal evidence of biased policing at the national level, I make several novel contributions. This study is the first correspondence study on local police departments and the first to focus on bias in the context of police accountability. A particularly significant element of this study is the inclusion of Hispanic and multiple gender identities. The majority of research concerning biased policing concerns itself with differences between White and Black citizens. However, as Weitzer (2014) points out, considering the growing population of Hispanic Americans, the lack of research on police-citizen interactions for Hispanics is "particularly puzzling." On its own, it is crucial to understand if police discriminate against a particular gender. Additionally, many stereotypes that may motivate biased policing frequently include the intersection of race/ethnicity and gender. By including race and gender in my study, I can test correlations between these problematic stereotypes and policing practices.

2 Experimental Design and Data

2.1 Experiment

In this section, I describe the correspondence study's design, creation, and implementation. The objective of the study is to test whether police departments exhibit signs of racial/ethnic or gender discrimination. The design of this study, in broad strokes, is to collect contact information for a sample of police departments and then contact the departments using identities that I created. I preregistered this experiment at the AEA RCT Registry, and the pre-analysis plan can be found here.

Police Department Selection: The police departments included in this study are a stratified random sample. For inclusion in the study, I required a police department to serve a local government (i.e., no state police) and serve a population of at least 7,500 people. To generate my sample, I randomly sampled the US Census's universe of local governments in batches of 1,000.[7] From the sample of local governments, I searched the internet for an email address for the corresponding police department; I conducted a unique search for each department (i.e., I did not use an LEA directory). Some governments did not have a local police department, and some police departments did not have publicly available email addresses. I recorded the issue in both cases and dropped the local government from the study. Many police departments had multiple publicly available email addresses. When deciding which address to select, I prioritized the general department email, then the police chief, and then the next-highest-in-command officer. In the end, I selected 2,134 departments to receive emails, representing 49 states.[8] Figure 1 illustrates the proportion of a state's population that is represented by the departments selected for this study. To calculate these proportions, I summed the local populations served by departments in a state and divided that sum by the state's total population. Please refer to appendix A for more details.

Identity Creation: I use the names of the email senders to signal race and gender. I created six broad categories of identities for this study: Black female, Black male, Hispanic female, Hispanic male, White female, and White male. Sixty unique first-last name combinations represent each identity.

[7] I filtered the universe of local governments to exclude states, counties, and all governments with populations less than 7,500.

[8] This study does not include any Hawaiian police departments. Hawaii's exclusion was an unintentional result of the sampling process.

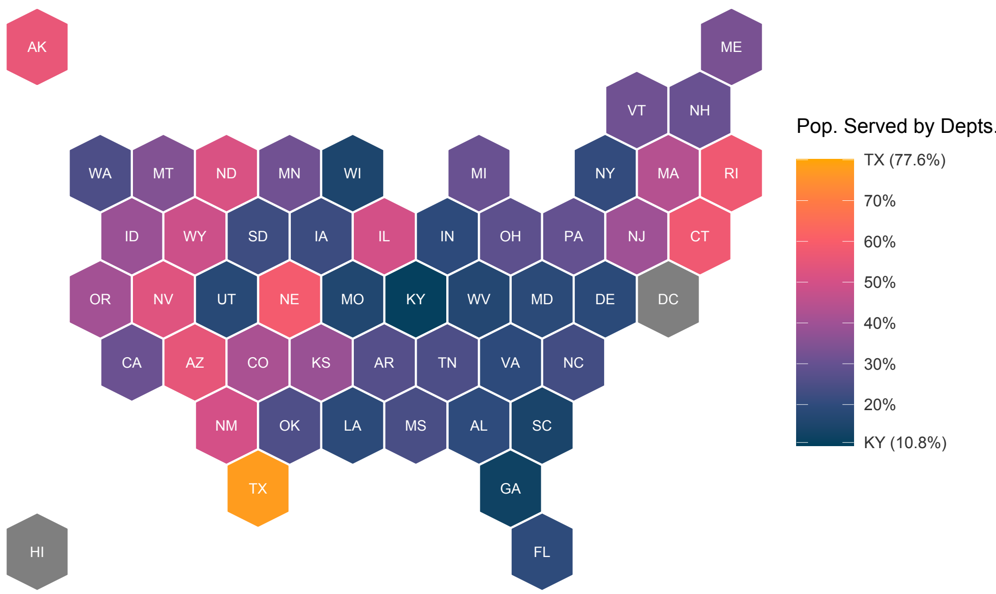


Figure 1: *The figure portrays the proportion of each state population represented by the departments I contacted. To calculate the proportion, I summed the local populations for each department in a state and then divided that sum by the state's total population. The state with the lowest representation was Kentucky (11%), and the state with the highest representation was Texas (77%)*

For this study, I selected last names from the "Frequently Occurring Surnames in the 2010 Census" dataset. I selected last names that were both racially distinctive and commonly occurring. To select racially distinctive names, I found names highly concentrated in one racial group—this requires that the name is both common for a particular race and uncommon for other races. However, in some cases, the most racially distinctive names were not commonly used in the United States. I avoided using very uncommon surnames to avoid the suspicions of the police departments. I constructed a simple search equation to find sufficiently racially distinctive *and* reasonably common names. I selected six last names for each race.

I referred to Gaddis (2017b) and Gaddis (2017a) to select the first names. Motivated by the frequent use of names as signals for races in audit studies, Gaddis conducts two experiments that explicitly test which first and last names are racially distinctive. In the experiments, Gaddis asked subjects which race they associated with a particular name. Gaddis conducted this experiment for names commonly used to represent Black people (Gaddis, 2017b) and Hispanic people (Gaddis, 2017a) in audit studies. I chose the ten most racially distinctive first names for the respective identities from these two studies. In total, I created 360 unique names (6 identities $\times$ 6 last names $\times$ 10 first names). After selecting the names for each identity, I created a unique email address for each last name used in the study (e.g., *olson.2922@mail.com*). I then created a unique email address profile for each identity (e.g., *Claire Olson* <*olson.2922@mail.com*>). As a result, the full names of the identities were visible in police departments' inboxes (e.g., *Claire Olson* <*olson.2922@mail.com*>).[9]

The complete list of names can be inferred from tables 1 and 2 (360 unique name combinations). I omitted six high-profile recognizable celebrity names: Denzel Washington, Tyra Banks, DaShawn Jackson, Seth Meyer(s), Katelyn Olson, and Pedro Martinez. These names have widespread recognition, and during the testing process, respondents noted that they strongly associate these names with celebrities having the same name.

Email: Each department received one email from a single randomly assigned identity. All emails were identical with two exceptions (1) the name of the email sender and (2) the sign-off used in the email. I varied the sign-off to test whether the tone of the email impacted police behavior. I decided to use the sign-off as

[9]Please refer to Appendix C for details concerning the specific email addresses used.

Table 1: Last names used

White	Black	Hispanic
Olson	Washington	Hernandez
Schmidt	Jefferson	Gonzalez
Meyer	Jackson	Rodriguez
Snyder	Joseph	Ramirez
Hansen	Williams	Martinez
Larson	Banks	Lopez

Table 2: First names used in study

White Male	White Female	Black Male	Black Female	Hispanic Male	Hispanic Female
Hunter	Katelyn	DaShawn	Tanisha	Alejandro	Mariana
Jake	Claire	Tremayne	Lakisha	Pedro	Guadalupe
Seth	Laurie	Jamal	Janae	Santiago	Isabella
Zachary	Stephanie	DaQuan	Tamika	Luis	Esmeralda
Todd	Abigail	DeAndre	Latoya	Esteban	Jimena
Matthew	Megan	Tyrone	Tyra	Pablo	Alejandra
Logan	Kristen	Keyshawn	Ebony	Rodrigo	Valeria
Ryan	Emily	Denzel	Denisha	Felipe	Lucia
Dustin	Sarah	Latrell	Taniya	Juan	Florencia
Brett	Molly	Jayvon	Heaven	Fernando	Juanita

the control because it influences the amicability of the email but minimally changes the content of the email. I randomly assigned the sign-off across emails. I use the email sender's name twice in each email to increase salience. Each email had the following format:

From: *Full name* <*lastname19xx*@mail.com>

Subject: Complaint Assistance

Body:

X Police Department,

My name is *first name* and I am interested in filing a complaint against an officer in your department. I am not sure what to do, and would like to request information on how to make a complaint. Can you please send me this information?

Sign off

Full name

The *italicized* words indicate that these words changed across emails. As seen above, I created profiles for the email accounts so that departments would see an identity's full name twice and an identity's first name thrice. Police departments were addressed directly—without, for example, a "Hello"—because I found in the testing process that the inclusion of a salutation increased the chances of the email being marked as spam. The sign-off varied between a cordial sentiment (*"Thank you!"*) and a curt sentiment (*"Sincerely,"*). Appendix D includes images of example emails and other details on the design of the email template.

Timing: I conducted the study over a ten-week period, from late June 2022 to late August 2022. I sent roughly 210 emails each week, split across Monday, Tuesday, and Wednesday. I sent all emails at approximately 9 a.m. local time for each police department. I rolled the experiment out over ten weeks to minimize the chance that a single event compromised the generalizability of the results. Splitting the emails across days of the week merely reduced the logistical difficulty of sending the emails. I did not send emails on Thursday, Friday, or weekend days to give departments at least two full weekdays to respond.

Treatment Assignment The "treatment" here is the identity (race and gender) that each department sees. I stratified treatment by week and by state. As a result, the number of departments for each state is balanced each week. Treatment was then randomly assigned across departments within each week-state stratum. Appendix E details the treatment assignment process.

2.2 Data

I use several additional datasets in this study. As mentioned, I used data from the US Census to create a pool of local governments in the department-selection process (U.S. Census Bureau (2021)). I limited the local governments eligible for inclusion in the study to exclude state and county governments and governments with populations less than 7,500 residents.[10] I then matched selected departments with police department directories from OpenPolice.org and ICPSR (Lesko et al., 2021; United States and Bureau of Justice Statistics, 2012)). Department directories added information about agency locations and unique identification numbers.[11]

The study includes data for several other observable department characteristics. These characteristics, ex-ante, seemed to be potentially important determinants of the response behaviors of police departments: numbers of officers and civilian employees for each department; county-level income information; and county-level racial/ethnic composition. I use UCR data codified by Kaplan (2021) for employee counts for each department. The UCR dataset includes employee counts through 2020. However, a handful of departments are missing data for 2020. I use the most recent employee count since 2010 (231 departments) where available. If a department does not have an employee count after 2010, I record that department's employee count as missing (29 departments). I take income and race data from the 2019 American Community Survey (U.S. Census Bureau, 2019). In the 1-year ACS data, 148 counties are missing data for the median income of Black households, and 55 counties are missing data for the median income of Hispanic households. Police departments selected for the study are associated with governments smaller than counties. However, it is unclear with exactly which population each department would interact. If I use data with too precise a

[10] During the collection process for police department email addresses, I accidentally included 117 departments in communities with populations less than 7,500. I included these departments in the study.

[11] I collected the *Originating Agency Identifier* (ORI) numbers for all departments that have ORI. See Office of Justice's explanation for details.

geography (e.g., the zip codes of the departments), I risk mischaracterizing a department's local conditions. Accordingly, I use county-level data to characterize the economic and racial composition of a department's local area. I sacrifice precision with this approach but avoid inaccuracy.

Table 3 shows relevant department characteristics. Column 2 of the table is the mean value for each different characteristic for departments that received emails from White-male identities. Columns 3 through 7 are the differences between the White-male mean value and the other identities. Table 3 confirms that the treatment was successfully randomized across the most obvious department characteristics relevant to this study. Only one of the 80 differences throughout the rows and columns is statistically significant, and only at the 10% level (*Pop. % Black (county-level)*).

Table 3: This table compares mean values of geographic and police departmental characteristics across each identity. Column 1 delineates which variable is being compared. Column 2 displays the mean value of the variables for departments that received emails from White male identities. Columns 3 through 7 show the difference between the value in Column 2 and the mean value for the other five identities. For example, the average county median income for departments that received emails from White male identities is $66,700. In comparison, the average county median income for departments that received emails from Black female identities is $100 lower, with a standard error of $1,300. The absence of statistical significance in the table reflects that the variables are not correlated to treatment assignment.

		Putative Identity				
	White Male (n = 359)	White Female (n = 352)	Hispanic Male (n = 350)	Hispanic Female (n = 361)	Black Male (n = 358)	Black Female (n = 354)
	(Mean)	*Differential*				
Income (county-level)						
Median Income all HH (hundreds of dollars)	$667	-0.3 *(13)*	0.7 *(13)*	-5 *(13)*	3.4 *(13)*	-1 *(13)*
Median Income Black HH (hundreds of dollars)	$477	-19 (15)	5 (15)	-1.7 (15)	6.3 (15)	7.6 (15)
Median Income Hispanic HH (hundreds of dollars)	$531	-6.8 (10)	3.3 (10)	-6.8 (10)	-2.7 (10)	4.8 (10)
Median Income White HH (hundreds of dollars)	$733	9.6 (14)	6.8 (15)	-12 (14)	3.4 (14)	3.5 (14)
% Pop. in poverty	12	0.39 (0.36)	0.40 (0.36)	0.12 (0.35)	-0.11 (0.35)	0.21 (0.36)
% Black pop. in poverty	22	0.90 (0.76)	-0.07 (0.77)	-0.20 (0.76)	-0.61 (0.76)	0.33 (0.76)
% Hispanic pop. in poverty	19	-0.04 (0.56)	0.25 (0.57)	0.13 (0.56)	0.00 > (0.56)	-0.16 (0.56)
% White pop. in poverty	9	0.06 (0.27)	0.08 (0.27)	-0.06 (0.27)	-0.17 (0.27)	0.02 (0.27)
Population						
Local government pop. (hundreds)	491	41 (128)	-11 (128)	130 (127)	-104 (127)	17 (127)
Pop. % Black (county-level)	10	**1.57* (0.84)**	0.19 (0.85)	0.64 (0.84)	0.07 (0.84)	0.84 (0.84)
Pop. % Hispanic (county-level)	14	-0.21 (1.11)	0.72 (1.11)	0.37 (1.10)	0.16 (1.11)	0.33 (1.11)
Pop. % White (county-level)	69	-1.99 (1.48)	-1.15 (1.48)	-1.32 (1.47)	-0.23 (1.47)	-1.67 (1.48)
Pop. % rural (county-level)	21	-2.19 (1.51)	-1.29 (1.51)	-0.69 (1.50)	0.61 (1.50)	-0.43 (1.51)
Department size (# of employees)						
Total employees	128	-1 (43)	-13 (43)	48 (43)	-22 (43)	-8 (43)
Total officers	104	-1 (36)	-12 (36)	40 (36)	-22 (36)	-9 (36)
Total civilian employees	24	0 (8)	-1 (8)	8 (8)	0 (8)	1 (8)

*Standard-errors in parentheses. Signif. Codes: *: 0.1*

Note: 117 departments served local populations < 7,500. Department size data was missing for 29 departments, and 2020 department size data was missing for an additional 231 departments. 148 observations were missing for median income of Black households and 55 observations were missing for median income for Hispanic households in each department's county.

3 Results

3.1 Summary Statistics

I sent the first emails on Monday, June 27, 2022, and delivered the last emails on Wednesday, August 31, 2022. In total, I attempted to contact 2,134 police departments. Table 4 summarizes the outcomes of these emails. My final analyses (below) exclude the 37 denied or failed emails.[12]

Table 4: Emails Categorized by Outcome

Email Outcome	**Total**	**Percent of Total**
Sent	2,134	100.00%
Response	1,413	**66.21%**
No Response	682	31.96%
Multiple Response	207	9.70%
Denied	15	0.70%
Failed	24	1.12%

I categorize the emails by the outcome. The results show an overall response rate of 66.2%. Thirty-nine emails were undelivered because the police department's address needed to be corrected (failed) or police departments blocked the emails (denied). The response is slightly higher (67.5%) if I drop the 39 undelivered emails from the calculation. Of the 1,413 departments that responded, 207 departments sent multiple emails.

Table 4 indicates a response rate of 66.21%. If I exclude the 37 undeliverable emails (*Denied* or *Failed*), the response rate is 67.45%. This response rate is aggregated across all identities and says nothing about

[12]During the experiment, I received "undeliverable" emails from the email server I used. These "undeliverable" emails explained why the email I sent could not be delivered. In some cases, the email address I used for a police department was incorrect or no longer existed; these are the "failed" emails. In other cases, the police department email server blocked my email; these are the "denied" emails.

biased policing. The *Denied* category in column 1 represents emails that police departments blocked. The small number of *Denied* emails does not cause concern for the experiment's validity. However, it is concerning that some police departments have structured their email server to block a seemingly legitimate request for help. Of course, the request is part of an experiment, but it is easy to imagine a citizen with a genuine complaint making a similarly formatted request. The *Failed* emails are also a cause of concern in terms of police accountability. Because I manually collected the email addresses for each police department from their website, a *Failed* email potentially reflects a department's neglect in maintaining updated and accessible contact information.

I summarize department response times in Figure 2. The large majority of responses from police departments occurred in the first 24 hours that I sent the email, and I received 97% of the responses within two days. The expediency of responses suggest that departments take the request for help seriously. Combining the two results paints a picture that only some departments are willing to assist in making a complaint against an officer, but the willing departments do so actively.[13]

As mentioned, emails were sent out in batches over ten weeks to reduce the chances of current events influencing police department response behavior. In Appendix H, Figure B2 depicts the response rate for all identities by week, and Figure B1 breaks the weekly response rates down by identity. The figures suggest that response behavior did not change considerably, at least during the ten weeks of the experiment.

3.2 Main results

The primary focus of this study concerns the effect of racial or gender biases on transparency in policing. To do this, I estimate variations of the following equation:

$$Response_i = \beta_1 \Pi(Gender_i = Female) + \beta_2 \Pi(Race_i = Black) + \beta_3 \Pi(Race_i = Hispanic) + FE_{s,t} + \varepsilon_i$$

Where i indexes individual police departments, t indexes the week the email is sent, s indexes a department's state, and r indexes the race of the identity. The emphasis of analysis examines the differences in police department response behaviors to White putative identities and Black/Hispanic putative identities. Accord-

[13] In Appendix G, I examine if response time differs systematically across identities but find no evidence.

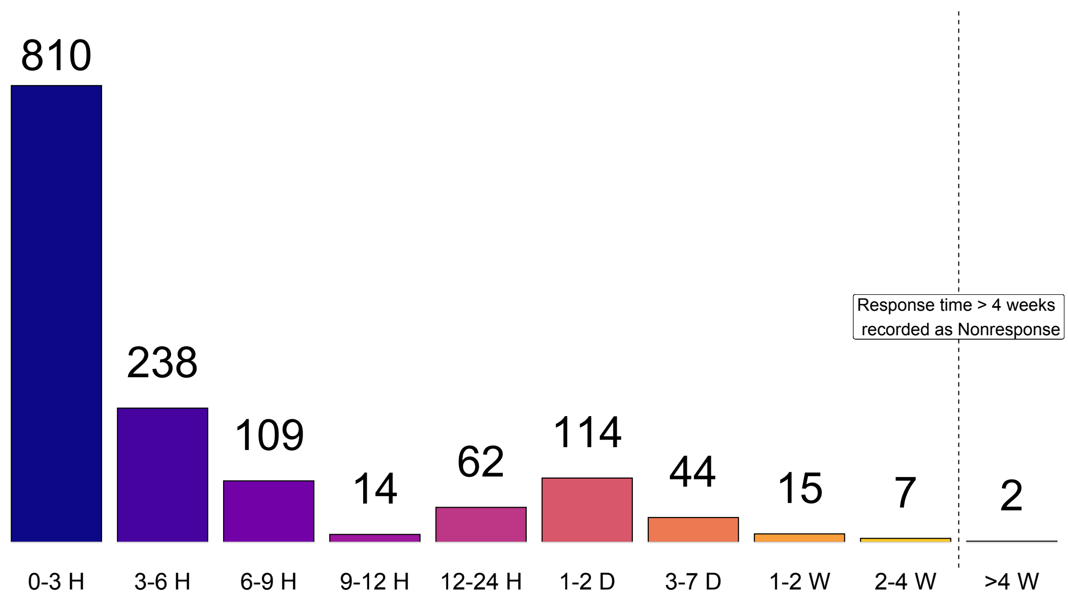


Figure 2: *I bin the number of responses by response time. I received most of the responses (809) within three hours of initial contact. Out of all the responses, I received 97% within 48 hours of contacting the departments. I received two emails outside the 4 week window, and I accordingly recorded these emails as non-responses.*

ingly, the omitted identity in the analysis is either White or White male. The main outcome, $Response_i$, for this study is a binary indicator for whether or not a police department responded, in any way at all, to an email. I record a response for a department if that department replies within four weeks (28 days).[14] FE represents fixed effects for the week I sent an email and a department's state. Because I stratified the treatment assignment on week and state, I include fixed effects for both throughout my analyses. Additionally, I cluster standard errors by week and state.

Table 5 reports the most general analysis of differences in response rates across identities using two weighting schemes. The alternative weighting schemes allow us to infer two different population parameters. The "unweighted" results describe response rates for an average police department. In other words, if a citizen were to contact a randomly selected police department, the unweighted response propensities (columns 2 and 4) are relevant. However, departments that serve larger populations interact with more citizens and are thus likely to receive more requests for assistance. By weighting each observation by that department's local population, the response rates shift the interpretation of the key coefficients from the average department's behavior to what the average citizen should expect to encounter. [15]

Column 1 of Table 5 compares unweighted department response rates for emails with Hispanic identities (Hispanic emails) and Black identities (Black emails) to the mean response rate for emails with White identities (White emails). The response rate for White emails is 74.86%. Compared to the White email response rate, the response rate for Black emails is 10.42 percentage points (pp) lower [4.33, 16.49], and the response rate for Hispanic emails is 10.66 pp lower [5.54, 15.77]. Both estimates are statistically significant at the 1% level.[16] Column 2 repeats the comparison in column 1 while weighting observations by the local population. Estimates across columns 1 and 2 are effectively identical for Black emails. Weighting by population marginally increases discrimination against Hispanic emails from 10.66 pp to 11.38 pp [8.48, 14.28].

[14]Automatically generated emails from departments only acknowledging that the treatment email was received I did not record as responses. I discuss alternative definitions of the response variable below.

[15]The square root of the population is used as a weight instead of simply the population because of the large distribution of population sizes. For example, Los Angeles has a population of close to 4 million, which is over 200 times as large as the median local population (18,000). However, using the standard method of logging the populations would reduce the disparity between populations too much. The log of Los Angeles's population is approximately 15, which is comparatively similar to the log of the median local population ($log(18000) \approx 9.8$).

[16]Comparison of the coefficients of the response rate for Black emails and the response rate for Hispanic emails reveals that the estimates are not statistically significant from each other.

Columns 3 and 4 compare department response rates for emails with female identities to the mean response rate for emails with male identities (66.03%). The unweighted estimate from column 3 shows that females, on average, were 2.26 pp more likely [-3.92, 8.44] to receive a response. However, the difference is not statistically significant. Column 4 indicates that when weighting by population, the response rate difference between females and males shrinks to 0.35 pp [-5.91, 6.60].

Table 5: Response Rate Differences by Race and Gender

Dependent Variable:	response			
Model:	(1)	(2)	(3)	(4)
Variables				
Black	-0.1042***	-0.1041***		
	(0.0310)	(0.0285)		
Hispanic	-0.1066***	-0.1138***		
	(0.0261)	(0.0148)		
Female			0.0226	0.0035
			(0.0315)	(0.0319)
Fixed-effects				
week	Yes	Yes	Yes	Yes
dept_address_state	Yes	Yes	Yes	Yes
Weights				
	Standard OLS	Sqrt of local pop.	Standard OLS	Sqrt of local pop.
Fit statistics				
Observations	2,095	2,095	2,095	2,095
R^2	0.05993	0.06897	0.04937	0.05700
Within R^2	0.01171	0.01271	0.00061	0.00001

Clustered (week & dept_address_state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

I compare differences in response rates for races and genders, unweighted and weighted by population. Black and Hispanic identities were less likely to receive responses from police departments than White identities. Black and Hispanic response rates were, respectively, 10.42 pp and 10.66 pp lower than the White response rate (74.82)—both significant at the 1% level. Females were marginally more likely, 2.23 pp, to receive responses than males (66.02).

3.2.1 Interacted Results

Literature shows that race and gender are often related to discrimination (Bertrand and Duflo, 2017). There is also evidence that the *intersection* of race and gender is an essential dimension of discrimination (e.g., Browne and Misra, 2003; Ifatunji and Harnois, 2016)). The intersectionality of race and gender also plays a significant role in the criminal justice system (e.g., Doerner and Demuth, 2010; Steffensmeier et al., 1998, 2017). Motivated by the significance of race-gender intersectionality in discrimination and the criminal justice system, I test whether this intersectionality plays a role in discriminatory policing. I do so by estimating the following equation:

$$
\begin{aligned}
Response_i = {} & \beta_1 White_i \times Female_i + \beta_2 Black_i \times Male_i + \beta_3 Black_i \times Female_i \\
& + \beta_4 Hispanic_i \times Male_i + \beta_5 Hispanic_i \times Female_i + FE_{s,t} + \varepsilon_i
\end{aligned}
$$

Where each β indicates the difference in response rate for each identity from the omitted White male response rate. I omit *White male* for two reasons. First, estimates compare the groups commonly discriminated against (non-White and female) to the group commonly given preferential treatment (White males). Second, this choice makes for the most straightforward interpretation of results as the *White male* identity has the highest response rate (75.78) among the six identities. Column 1 of Table 6 compares response rates across the six different identities (3 race categories $\times$ 2 gender categories). Column 2 reports the results of the same estimation equation as column 1 but weights observations by the local populations of the police departments.

Column 1 gives the percentage-point differential in response rates for the five identities compared to the White male identity. Column 2 reveals that response rates for Black and Hispanic males were significantly lower than White males at the 1% level and are the lowest among all the identities. Specifically, Black males were 13.94 [6.55, 21.33] pp less likely to receive a response, and Hispanic males were 15.00 [8.05, 21.94] pp less likely to receive a response than White males. The corresponding response rates for females (specifically Black and Hispanic) were higher than their male counterparts but still significantly lower than White males. Black females were 9.70 [1.92, 17.48] pp less likely to receive a response and Hispanic females were 9.28 [0.99, 17.58] less likely to receive a response. The estimates are statistically significant at the 5% and 10% levels.

Testing for equality between the coefficients within each race group between genders finds that the response rates for Black males and Black females are not statistically significant (p-value = 0.3119). In contrast, the response rates for Hispanic males and females are statistically significant (p-value = 0.0035).

Table 6: Response Rate Differences for Race and Gender Interactions

Dependent Variable:	Response	
Model:	(1)	(2)
Variables		
White × Female	-0.0285	-0.0650
	(0.0479)	(0.0550)
Hispanic × Male	-0.1500***	-0.1465***
	(0.0354)	(0.0408)
Hispanic × Female	-0.0928*	-0.1391**
	(0.0423)	(0.0515)
Black × Male	-0.1394***	-0.1671***
	(0.0377)	(0.0459)
Black × Female	-0.0970**	-0.0996***
	(0.0397)	(0.0243)
Reference group mean		
White male	0.7578	
Weights		
	Standard OLS	Sqrt of local pop.
Fixed-effects		
Week	Yes	Yes
State	Yes	Yes
Fit statistics		
Observations	2,095	1,979
R^2	0.06214	0.07374
Within R^2	0.01404	0.01526

Clustered (Week & State) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Results display the difference in response rate for each identity compared to White males. The **White male response rate of 75.78** is the highest, and I use it as the control group. Model 2 has the exact specification as Model 1 but weights observations by the size of the local population for each department. Black and Hispanic males were the least likely to receive responses from police departments. Black and Hispanic females also have lower response rates than White males, but the magnitude and statistical significance of the estimates are not as large as their male counterparts. White females are marginally less likely to receive responses than White males, but the differences are not statistically significant.

The response rate for White females is 2.85 [-6.55, 12.25] percentage points lower but is not statistically significant at the 10% level. White females are the only female identity with a lower response rate than the male complement within each race/ethnicity grouping. The heterogeneous differences in response rates for gender, when interacted with race, suggest the importance of the intersectionality of race and gender. When individuals are White, males receive preferential treatment. However, this relationship inverts when the individual is Black or Hispanic. Likely, studying discrimination of race or gender without consideration of the other characteristic obscures the underlying situation.

Weighting by local populations increases the disparity in the response rate for White males and all the other identities, except for Hispanic males. The response rate for White females in column 2 (6.50 [-4.27, 17.28]) is more than double the estimate from column 1 but is still not statistically significant. When I include population weights, the response rates for Black females (9.96 [5.20, 14.73]) and Hispanic females (13.91 [3.81, 24.01]) increase in magnitude and statistical significance. The response rate for Black males (16.71 [7.72, 25.70]) becomes the identity with the lowest response rate. The response rate for Hispanic males (14.65 [6.66, 22.64]) decreases marginally but remains statistically significant at the 1% level. In contrast to column 1 of Table 6, testing for equality between the coefficients within each race/ethnicity group between genders finds that the response rates for Black males and Black females *are* statistically significant at the 10% level (p-value = 0.0861). The response rates for Hispanic males and Hispanic females *are not* statistically significant (p-value = 0.8545).

3.2.2 Department Size

Police department response rates likely depend on many factors. Department size and the population that a department serves could affect response rates from departments. For instance, large departments could staff personnel solely responsible for replying to requests for help making complaints. Conversely, smaller departments might be more sensitive to officer complaints if they are familiar with all officers in the department. Larger populations could mean that departments have more requests to fulfill. On the other hand, if departments serve small populations, they might be more suspicious of the genuineness of the email they receive.[17]

Table 7 displays the results of models with the same specifications as Table 6 when I split the data into “smaller departments” and “bigger departments”. I use the median number of total employees for the departments included in the study to determine a department’s size. The results in column 1 and column 2 reveal that smaller departments discriminate less against White females, Black males, and Black females than bigger departments. White females actually see a marginally higher response rate (1.62 [-11.84; 15.08]) than White males when interacting with smaller departments. Black male and female response rates are lower than White males for bigger and smaller departments. However, the response rate differentials are larger when interacting with bigger departments. For Black males, the response rate differential is more than twice as large for bigger departments (18.92 [1.59, 36.25]) than for smaller departments (8.92 [1.20, 16.64]). In contrast to White females, Black females, and Black males, Hispanic identities have higher response rates when interacting with bigger departments. Column 3 displays the results of explicitly testing for differences in identity response rate estimates across sample sizes. Of all the identities, only the Hispanic male response rate is significantly different. The lack of statistical significance could result from noisier estimates, as reflected by the larger standard errors resulting from halving the sample size.[18]

3.2.3 Department Email Type

As outlined in the experimental design section, there were inconsistencies in the public availability of police department email addresses. Consequently, the email address “type” I collected varied across departments. Here, I use “type” to refer to who/what is associated with the email address. For example, the email address I collected could be the general police department address or the email address of a police department’s chief.[19] In this study, I designate a police department as the unit of observation and use the email address I collected for each department as a proxy for the entire department. I examine the implications of this decision in the discussion section. In this section, I examine the relationship between the email address type and response rates across identities.

[17] Several responses mentioned that the police department checked the police logs and had no record of an interaction with a person that matched the name in the email they received.

[18] Please refer to Appendix I for a deeper analysis of the effect of department size on response rates.

[19] In Appendix A, I explain in detail how I collected the email addresses.

Table 7: Department Size Heterogeneity

Dependent Variable:	Response		
	Department Size		
Model:	Smaller (1)	Bigger (2)	**P value**
Variables			
White × Male mean	0.747	0.765	
White × Female	0.0162	-0.0703	0.5605
	(0.0595)	(0.0649)	
Hispanic × Male	-0.1755***	-0.1308**	> 0.0000
	(0.0401)	(0.0541)	
Hispanic × Female	-0.0892*	-0.0730	0.1183
	(0.0469)	(0.0605)	
Black × Male	-0.0892**	-0.1892**	0.6592
	(0.0341)	(0.0766)	
Black × Female	-0.0882	-0.1039	0.1611
	(0.0496)	(0.0569)	
Weights			
Sqrt of local pop.	No	No	
Fixed-effects			
Week	Yes	Yes	
State	Yes	Yes	
Fit statistics			
Observations	1,052	1,014	
R^2	0.08835	0.09749	
Within R^2	0.01864	0.01648	

Clustered (Week & State) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes. I split data into two samples according to the median department size. I compare response rates for each identity to the White male response rate for both samples. Black females, Black males, and White females have comparatively lower response rates when interacting with "bigger" departments. In contrast, Hispanic females and males have comparatively higher response rates when interacting with "bigger" departments. The response rate for White males is marginally higher when interacting with "bigger" departments. The only statistically different response rate across department sizes is for Hispanic males.

I categorized the email types I collected into six groups: *Department*, *Chief*, *Commanding Officer*, *Found*, *Records*, and *Accountability*.[20] As can be seen in Table 8, the vast majority of the emails collected are either *Department* or *Chief*. For this reason, I focus this analysis on these two email types.

Table 8: Email Type Overview

Email Category	**Count**	**Percent of Total (2095)**	**Response Rate**
Department	853	40.72%	0.659
Chief	1063	50.74 %	0.687
Commanding Officer	84	4.01%	0.726
Found	40	1.91%	0.45
Records	37	1.77%	0.865
Accountability	18	0.85%	0.556

Different "types" of email addresses were collected, out of necessity, for different police departments. Column 1, Email Category, describes the type of email address collected (definitions above). The majority of the email addresses used in this study were *Department* or *Chief* email addresses. Although there is variation in Column 4, the number of observations are quite low.

Because I did not stratify treatment assignment by email type, I test if there is a relationship between email types and treatment assignment. To do this, I estimate the following equation:

$$\begin{aligned} Email-Type_i = \beta_1 White_i \times Female_i + \beta_2 Black_i \times Male_i + \beta_3 Black_i \times Female_i \\ + \beta_4 Hispanic_i \times Male_i + \beta_5 Hispanic_i \times Female_i + \varepsilon_i \end{aligned}$$

Similarly to earlier analyses, I omit *White Male* and use it as the base rate. The results in Table 9

[20] Explanation of each of these groups can be found in Appendix A.

demonstrate that due in large part to the random assignment of treatment, there are minimal differences in the proportions of email types used for each identity. Column 1 of Table 9 describes the relationship between identity assignment and a police department's email type being *Department*. Column 2 of Table 9 describes the relationship between identity assignment and a police department's email type being *Chief*. Only the proportion of Black male emails sent to police chiefs (column 2) is statistically significant. About 47% of departments that received emails from White male identities had a *Chief* email type. Black male identities are 8.55 [1.15, 15.94] pp more likely to be assigned to departments with *Chief* email types than White male identities.

Next I compare response rates for identities between departments with *Department* email types and departments with *Chief* email types. To do so, I estimate the following equation twice:

$$\begin{aligned} Response_i = \beta_1 White_i \times Female_i + \beta_2 Black_i \times Male_i + \beta_3 Black_i \times Female_i \\ + \beta_4 Hispanic_i \times Male_i + \beta_5 Hispanic_i \times Female_i + FE_{s,t} + \varepsilon_i \end{aligned}$$

First, I estimate the equation while restricting the data to the *Department* email type and then the *Chief* email type. Table 10 reports the results. The second column of Table 10, *Baseline*, uses all of the data from the study (n = 2095) and is used as a point of reference for the other two models.

The estimates in the third and fourth columns of Table 10 indicate that for some identities, response rates can vary considerably depending on the email type of the police department. Except for Hispanic males, all identities had higher response rates when the department email type was *Chief*. The most striking result is for Black male identities. Police departments with a *Department* email address type were substantially more likely to discriminate against Black male identities than police departments that had a *Chief* email address type. When contacting a *Department* email address type, the Black male identities response rate was 22.03[9.96, 34.09] pp lower than White males—significant at the 1% level. However, Black male response rates were 8.12 [-1.11, 17.35] pp lower than White males when contacting *Chief* email types. While still lower than White males, the magnitude of discrimination is roughly 60% smaller and not statistically significant. Heterogeneous White male response rates across email types may drive a small portion of the

Table 9: Email Department Type Distribution Across Identities

Dependent Variables:	Department	Chief
Model:	(1)	(2)
Putative Identities		
White × Male	0.4188***	0.4701***
	(0.0262)	(0.0267)
Differential from White Male		
White × Female	-0.0304	0.0488
	(0.0373)	(0.0379)
Hispanic × Male	0.0156	0.0285
	(0.0373)	(0.0379)
Hispanic × Female	-0.0323	0.0509
	(0.0369)	(0.0376)
Black × Male	-0.0342	0.0855**
	(0.0371)	(0.0377)
Black × Female	0.0122	0.0098
	(0.0372)	(0.0378)
Fit statistics		
Observations	2,095	2,095
R^2	0.00188	0.00325
Adjusted R^2	-0.00051	0.00087

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

What is the relationship between identity assigned to a department and the email address type collected from the department? The first row *White X Male* describes the proportion of departments receiving an email from a White male identity based on the type of email type collected from that department. These values reflect the proportions of email types of emails collected (seen in Table 8). The subsequent rows show the differential for each identity from the baseline of White male.

Table 10: Response Rate by Department Email Type

Dependent Variable:	Response				
				Control for Dept. Size	
Model:	Baseline	Department	Chief	Department	Chief
Putative Identities					
White × Male *(mean)*	0.7580	0.7619	0.7515	0.7619	0.7515
White × Female	-0.0285	-0.0289	-0.0064	-0.0297	-0.0041
	(0.0479)	(0.0626)	(0.0714)	(0.0648)	(0.0723)
Hispanic × Male	-0.1500***	-0.1378**	-0.1438**	-0.1338**	-0.1372**
	(0.0354)	(0.0557)	(0.0594)	(0.0550)	(0.0604)
Hispanic × Female	-0.0928*	-0.1337*	-0.0644	-0.1268*	-0.0569
	(0.0423)	(0.0721)	(0.0539)	(0.0700)	(0.0513)
Black × Male	-0.1394***	-0.2203***	-0.0812	-0.2150***	-0.0877*
	(0.0377)	(0.0615)	(0.0471)	(0.0619)	(0.0473)
Black × Female	-0.0970**	-0.1073**	-0.0927	-0.1081**	-0.0915
	(0.0397)	(0.0408)	(0.0627)	(0.0409)	(0.0611)
Fixed-effects					
Week	Yes	Yes	Yes	Yes	Yes
State	Yes	Yes	Yes	Yes	Yes
Agency Size	No	No	No	Yes	Yes
Fit statistics					
Observations	2,095	853	1,063	838	1,051
R^2	0.06214	0.10994	0.11299	0.11252	0.11392
Within R^2	0.01404	0.02433	0.01212	0.02282	0.01177

Clustered (Week & State) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Different types of email addresses were collected for departments. In addition to heterogeneity of response rates across identities, there is additional heterogeneity across email types. Column 2, **Baseline**, are the response rate differences for each identity compared to the response rate for White male identities ignoring department email address type. Column 3 and 4 compare response rates across department email type. Column 5 and 6 mirror 3 and 4 but include a control for agency size.

differential. Police departments with a *Department* email address type responded to 76.19% of the White male emails, which was slightly higher than police departments with a *Chief* email address type (75.15%). Nonetheless, this does not explain such a dramatic difference.

A possible explanation for *Department* email types leading to more discrimination is that email type is a proxy for department size. There is a correlation between department size and which email address type I collected for a department. The bigger departments are more likely to have a departmental email address publicly available than departments that serve smaller populations. Consequently, police departments I collected a *Department* email type were significantly larger in personnel size than police departments for which I collected a *Chief* email type. Table 7 shows that response rates vary by department size. In the case of Black male identities, bigger departments were less likely to respond than smaller departments. This heterogeneity is consistent with departments with a *Department* email type responding less often to Black male identities.

I reestimate the models while controlling for department size and report the results in columns 5 and 6. A comparison of the results reveals that department size does not explain the significant difference in response rates across email types. There are slight differences, but the sample sizes are slightly smaller when controlling for department size. Instead, it appears that email type impacts response rates. One possible explanation is that departments that designate their police chief as the public point of contact differ from those that do not. Confounding this explanation is that many departments with a publicly available departmental email address also had an email address for the police chief. A more compelling explanation is that a police chief is unlikely to be the officer mentioned in the eventual complaint and is most interested in ensuring that subordinates maintain a high level of professionalism. In contrast, lower-ranking personnel likely maintain the departmental email address. These types of employees may be concerned that a complaint pertains to them. A department email may also offer anonymity not afforded to a police chief's email.

4 Discussion

4.1 Interpreting Bias

The results of this field experiment present substantial evidence of racially biased police practices. When aggregated across genders, compared to White emails, the response rates for Hispanic and Black emails are both 10 percentage points lower and statistically significant at the 1% level (Table 5). The question of gender-biased policing is more nuanced than racially biased policing. When comparing the pooled response rates for males to the pooled response rates for females, the results suggest that police departments are slightly more likely to respond to female requests. However, comparing response rates for each identity tells a different story (Table 6). White-male identities receive the highest rate of responses—a reversal of the female identities receiving preferential treatment result. The low response rates for Hispanic and Black males drive the lower response rates for all males. Hispanic and Black female response rates are 9 percentage points lower than White males and significant at the 10% and 5% levels, respectively. Comparing the results of Table 5 to Table 6 reveals that the *intersection* of race/ethnicity and gender is an essential part of the story. Hispanic males receiving the lowest response rates of all groups indicates the importance of expanding police-citizen relationship research to include Hispanic demographics (Weitzer (2014)).

Identifying the mechanism(s) for the hierarchy of response rates of the six identities is beyond the scope of this paper. However, it is worth considering why Hispanic *males* and Black *males* received the lowest rate of responses, despite White *males* receiving the highest rate of responses. The discrepancy could be explained by the historical narrative of black and brown males being viewed as criminals (e.g., the racist stereotype of the "superpredator"). A common rebuttal to this hypothesis is that these groups might be more likely to participate in crime—echoing the challenge researchers run into of separating biased policing from different levels of participation in criminal activities amongst different ethnic/racial groups. However, in the context of the present study, no crime has been committed. Black and Hispanic males are simply not treated the same way as their White counterparts.

An alternative explanation for the mechanism behind this discrepancy is that departments hypothesize that the nature of the complaint might differ across groups. For instance, research suggests that police are more likely to use force with people of color (e.g., Edwards et al., 2019; Fryer, 2020; Nix et al., 2017; Ross, 2015). The lower response rates from the present study might reflect that departments think that complaints from Hispanic and Black males are more likely to concern excessive use of force from one of their officers, a type of complaint that is more damaging to the department.

Previous work has suggested that documented racial/ethnic discrimination may reflect bias against poorer or less-educated communities—and race/ethnicity serves as a proxy for wealth and education (**?**). For instance, Giulietti et al. (2019) take measures to separate the two in their correspondence study. However, in practice, this distinction may not matter. The lived experience of Black and Hispanic populations includes bias—regardless of whether the bias results from racism or classism. It is likely that police who disproportionately target non-White groups are engaged to some degree in both statistical targeting and biased policing (Bertrand and Duflo, 2017). **?** argues that statistical discrimination "can lead people to view social stereotyping as useful and acceptable and thus help rationalize and justify discriminatory decisions."

4.2 Accountability

The primary question of this study seeks to answer whether police departments discriminate on race, ethnicity, or gender. The study's design also emphasizes another critical topic for policymakers interested in reforming police practices: accountability. The correspondence study's design forces police departments to decide to respond to an inquiry based on a citizen's race/ethnicity and gender. However, the study also asses the willingness of departments to assist a citizen attempting to hold one of their officers accountable. In the existing literature, the only prior correspondence study that includes law enforcement agencies is Giulietti et al. (2019). Their correspondence study interacts with many public institutions, including sheriff's offices. In their study, the authors email the various public institutions with benign requests for relevant information. They use two black male names and two white male names to vary the identity of the citizen asking for information. The authors find that these public institutions (ranging from public libraries to county clerks, in addition to sheriff's offices) are less likely to respond to emails from individuals with distinctively black names. Giulietti et al. (2019) find that response rates for their sheriff's offices are approximately 53% for White male emails and 46% for Black male emails. Overall, these response rates are noticeably lower than the average response rates in the present study. However, the difference in response rates by race is significantly smaller. One explanation for the difference is that Giulietti et al. (2019) targets sheriff's offices instead of local police departments, and sheriff's offices may face different expectations for accountability. However, an alternative explanation is that departments do not treat a simple request for general assistance with the same urgency as a request for help in making a complaint against a police officer. When a request for assistance concerns making a complaint, police departments appear more responsive but may be more likely to discriminate.

It should be noted that in both studies, Giulietti et al. (2019) and the present study, average response rates are low. The average response rate of 67.4% for the present study (with a low of 60.6% for Hispanic males) is concerning. Even the most responded to identity, White males, have a response rate of 75%. By design, the complaints mentioned in the present study are fictitious. However, in reality, a citizen attempting to file a formal complaint suggests potentially serious misconduct on the officer's part. Suppose only six out of ten citizens can obtain assistance making a complaint. In that case, citizen-initiated complaints about police officers may not present a reliable or just strategy for holding police officers accountable. This concern is amplified when groups of people who more often interact with police (i.e., people of color) are also less likely to be assisted in making a complaint.

4.3 Caveats

This study seeks to understand whether police departments tend to discriminate based on race, ethnicity, or gender. The results suggest that the average police department does. There are a few caveats to this study. First, police departments were (see Appendix A) selected randomly. However, a department was only eligible for inclusion in the study if it had a publicly available email address. There are likely to be non-random department characteristics that distinguish between departments that make their email addresses available to the public and those that do not. Consequently, this study's results reflect average department behavior only for a specific type of department. It is plausible that departments willing to share a contact email might also be more willing to engage with the public. Thirty-nine departments included in the study had contact emails found somewhere other than on the police department website (e.g., the police chief's contact information is posted on the city's website but not the police department's own website or the department's specific page on the city website). The response rate for emails found this way was almost 20 percentage points lower than the overall mean (47% versus 66%). Drawing clear inferences from such a small sample is challenging.

However, this difference in response rates suggests that departments with easier-to-find email addresses may be more willing to engage with the public.

Second, this analysis does not seek to identify a fixed effect for each police department. The results demonstrate that police departments in the United States have a higher propensity, on average, to respond to White emails than Black or Hispanic ones. However, the data provide only one observation per department. Thus it is not possible to infer systemic bias within individual police departments. Revisiting this type of RCT with a specific aim of learning more about within-department behavior may be of value in future studies.[21] Finally, one must keep in mind that this conclusion pertains to a specific context. This study demonstrates that police departments discriminate on race, ethnicity, or gender when contacted via email for help making a complaint against an officer. It is unclear the extent to which the bias detected in this study is present in other contexts, for example, a police officer's decision to pull over a vehicle.

4.4 Conclusion

This study uses a correspondence study to establish strong causal evidence of biased policing in the United States. Across 2,134 police departments, departments were 10 percentage points more likely to respond to emails from White identities than Black or Hispanic identities. Interacting the race/ethnicity and gender of the identities revealed that White male identities had the highest response rates and Black male and Hispanic male identities had the lowest response rates—respectively, 13.94 and 15 percentage points lower than the White male response rates. The low overall response rates and significant bias in responses across identities are each concerning. Low response rates suggest police departments resist accountability. Bias in responding to minority identities suggests that departments are especially unwilling to engage with communities of color—disproportionately policed communities. While the existing literature has been inconclusive about the existence of biased policing, the results of this study suggest that bias in policing does exist and that it may hinder progress toward police transparency and accountability.

[21] Multiple requests to one department may raise suspicions about the inquiries.

References

Antonovics, K. and Knight, B. G. (2009). A new Look at racial profiling: Evidence from the Boston Police Department. *Review of Economics and Statistics*, 91(1):163–177.

Ba, B. (2016). How Far Are You Willing to Go against the Police? Evaluating the Effects of Citizen Affidavits in Chicago.

Bertrand, M. and Duflo, E. (2017). Chapter 8 - Field Experiments on Discrimination. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments*, volume 1, pages 309–393. North-Holland.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.

Browne, I. and Misra, J. (2003). The Intersection of Gender and Race in the Labor Market. *Annual Review of Sociology*, 29:487–513.

Bulman, G. (2019). LAW ENFORCEMENT LEADERS AND THE RACIAL COMPOSITION OF ARRESTS. *Economic Inquiry*, 57(4):1842–1858. Publisher: Blackwell Publishing Inc.

Butler, D. M. and Broockman, D. E. (2011). Do politicians racially discriminate against constituents? A field experiment on state legislators. *American Journal of Political Science*, 55(3):463–477.

Chalfin, A., Hansen, B., Weisburst, E., and Williams, M. (2021). Police Force Size and Civilian Race. *SSRN Electronic Journal*.

Chalfin, A. and McCrary, J. (2018). Are U.S. cities underpoliced? Theory and evidence. *Review of Economics and Statistics*, 100(1):167–186.

Chen, M. K., Christensen, K. L., John, E., Owens, E., and Zhuo, Y. (2021). Smartphone Data Reveal Neighborhood-Level Racial Disparities in Police Presence.

DeAngelo, G. and Hansen, B. (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy*, 6(2):231–257.

Doerner, J. K. and Demuth, S. (2010). The independent and joint effects of race/ethnicity, gender, and age on sentencing outcomes in U.S. federal courts. *Justice Quarterly*, 27(1):1–27.

Edwards, F., Lee, H., and Esposito, M. (2019). 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 of the United States of America*, 116(34):16793–16798. Publisher: National Academy of Sciences.

Einstein, K. L. and Glick, D. M. (2017). Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats and Access to Public Housing. *American Journal of Political Science*, 61(1):100–116. Publisher: Blackwell Publishing Ltd.

Fridell, L. A. (2017). Explaining the Disparity in Results Across Studies Assessing Racial Disparity in Police Use of Force: A Research Note. *American Journal of Criminal Justice*, 42(3):502–513. Publisher: Springer New York LLC.

Fryer, R. G. (2020). An empirical analysis of racial differences in police use of force: a response. *Journal of Political Economy*, 128(10):4003–4008.

Gaddis, S. M. (2017a). how black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. *Sociological Science*, 4:469–489. Publisher: Society for Sociological Science.

Gaddis, S. M. (2017b). Racial/Ethnic Perceptions from Hispanic Names: Selecting Names to Test for Discrimination. *Socius: Sociological Research for a Dynamic World*, 3:237802311773719. Publisher: SAGE PublicationsSage CA: Los Angeles, CA.

Gelman, A., Fagan, J., and Kiss, A. (2007). An analysis of the New York City police department's "stop-and- frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, 102(479):813–823.

Giulietti, C., Tonin, M., and Vlassopoulos, M. (2019). Racial discrimination in local public services: A field experiment in the United States. *Journal of the European Economic Association*, 17(1):165–204.

Goncalves, F. and Mello, S. (2021). A few bad apples? racial bias in policing. *American Economic Review*, 111(5):1406–1441. Publisher: American Economic Association.

Ifatunji, M. A. and Harnois, C. E. (2016). An Explanation for the Gender Gap in Perceptions of Discrimination among African Americans: Considering the Role of Gender Bias in Measurement. *Sociology of Race and Ethnicity*, 2(3):263–288. Publisher: SAGE Publications Inc.

Kaplan, J. (2021). Jacob Kaplan's Concatenated Files: Uniform Crime Reporting Program Data: Law Enforcement Officers Killed and Assaulted (LEOKA) 1960-2020.

Knox, D., Lowe, W., and Mummolo, J. (2020). Administrative Records Mask Racially Biased Policing. *American Political Science Review*, 114(3):619–637.

Lesko, M., Silverman, S., and Troup, C. (2021). Police Departments by State and County.

Luh, E. (2019). Not So Black and White: Uncovering Racial Bias from Systematically Masked Police Reports. *SSRN Electronic Journal*.

Makowsky, M. D., Stratmann, T., and Tabarrok, A. (2019). To serve and collect: The fiscal and racial determinants of law enforcement. *Journal of Legal Studies*, 48(1):189–216. Publisher: University of Chicago Press.

Mello, S. (2019). More COPS, less crime. *Journal of Public Economics*, 172:174–200. Publisher: Elsevier B.V.

Nix, J., Campbell, B. A., Byers, E. H., and Alpert, G. P. (2017). A Bird's Eye View of Civilians Killed by Police in 2015: Further Evidence of Implicit Bias. *Criminology and Public Policy*, 16(1):309–340.

Oberfield, Z. W. and Incantalupo, M. B. (2021). Racial Discrimination and Street-Level Managers: Performance, Publicness, and Group Bias. *Public Administration Review*, 81(6):1055–1070. Publisher: John Wiley and Sons Inc.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., and Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, 4(7):736–745. arXiv: 1706.05678 Publisher: Nature Publishing Group.

Ross, C. T. (2015). A multi-level Bayesian analysis of racial Bias in police shootings at the county-level in the United States, 2011-2014. *PLoS ONE*, 10(11):e0141854. Publisher: Public Library of Science.

Ross, C. T., Winterhalder, B., and McElreath, R. (2018). Resolution of apparent paradoxes in the race-specific frequency of use-of-force by police. *Palgrave Communications*, 4(1):1–9. Publisher: Palgrave.

Sances, M. W. and You, H. Y. (2017). Who pays for government? descriptive representationb and exploitative revenue sources. *Journal of Politics*, 79(3):1090–1094.

Smith, M. R., Rojek, J. J., Petrocelli, M., and Withrow, B. (2017). Measuring disparities in police activities: a state of the art review. *Policing*, 40(2):166–183.

Steffensmeier, D., Painter-Davis, N., and Ulmer, J. (2017). Intersectionality of Race, Ethnicity, Gender, and Age on Criminal Punishment. *Sociological Perspectives*, 60(4):810–833. Publisher: SAGE Publications Inc.

Steffensmeier, D., Ulmer, J., and Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology*, 36(4):763–798. Publisher: American Society of Criminology.

Stroube, B. K. (2021). Using allegations to understand selection bias in organizations: Misconduct in the Chicago Police Department. *Organizational Behavior and Human Decision Processes*, 166:149–165.

United States and Bureau of Justice Statistics (2012). Law Enforcement Agency Identifiers Crosswalk, United States.

U.S. Census Bureau (2019). Selected Housing Characteristics 20105-2019 American Community Survey 5-year estimates.

U.S. Census Bureau (2021). U.S. Population Estimates 2010-2020.

Weisburd, S. (2021). Police presence, rapid response rates, and crime prevention. *Review of Economics and Statistics*, 103(2):280–293. Publisher: MIT Press Journals.

Weisburst, E. K. (2019a). Police Use of Force as an Extension of Arrests: Examining Disparities across Civilian and Officer Race. *AEA Papers and Proceedings*, 109:152–156.

Weisburst, E. K. (2019b). Safety in police numbers: Evidence of police effectiveness from federal cops grant applications. *American Law and Economics Review*, 21(1):81–109. Publisher: Oxford University Press.

Weitzer, R. (2014). The puzzling neglect of Hispanic Americans in research on police–citizen relations. *Ethnic and Racial Studies*, 37(11):1995–2013. Publisher: Informa UK Limited.

West, J. (2018). Racial Bias in Police Investigations. (October):1–36.

White, A. R., Nathan, N. L., and Faller, J. K. (2015). What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. *American Political Science Review*, 109(1):129–142.

A Online Appendix: Police Department Selection

The selection process for police departments to be included in the study is as follows:

1. From the universe of governments provided by the U.S. Census, I create a list of possible jurisdictions that may have their own police departments. This list excludes state governments, counties, special districts and places with populations less than 7,500.
2. From the "possible department" list, I randomly a draw 1,000 jurisdiction names.[22]
3. Email addresses are then collected from police department websites in these jurisdictions.[23]
   - Governments without local police departments are dropped.
   - Police departments without publicly available email addresses are documented and dropped.
   - In the cases where there are multiple email addresses the prioritization is given first to (1) the email address for the department in general, and then to (2) the email address specifically for the police chief, (3) and finally to any possible contact (e.g., a community-affairs officer). I document the type of email address ultimately recorded in my database.
4. I repeat Steps 2 and 3 until 2,000 email addresses have been collected.[24]

Randomization of the department selection process increases the external validity of the study. Requiring that populations served by these police departments are greater than 7,500 increases the plausibility of the existence of the purported email sender as a resident of their jurisdiction.

Include a visual of where the addresses came from

Summary stats for dropped etc departments

A.1 Type of email collected

During the police department email address collection, the "type" of email address publicly available for collection varied from department to department. In this case, "type" refers to who is associated with the

[22] The target number of departments is 2,000, but to streamline the process, I select possible department cities in 3 batches of 1,000. For each batch of jurisdictions, roughly 60% have a viable police department, police chief or alternate email addresses. For a full explanation of this process see **??**

[23] For an example of how this process works please see **??**

[24] Given that jurisdictions are selected in batches of 1,000, the final number of police department emails collected is 2,135.

email address. For example, for Department **X** the only publicly available email address is for the chief of that department and for Department **Y** the only publicly available email address is for the shift-commander. In this example, I collect the email address for each department, and record that the *email address type* for Department **X** is *chief* and the *email address type* for Department **Y** is *shift-commander*. During the actual department email address collection, frequently departments had multiple email addresses publicly available.[25] In the case of the existence of multiple publicly available email addresses, I used a consistent priority list to decide which email address to collect. Prioritization is as follows:

1. Top priority is given to a general department email address. This is done to get the most accurate representation of a department's general behavior.

2. In the absence of a general department email address, priority is given to the chief of police.

3. In the absence of a general department email address or a chief email address, priority is given to the next highest in command officer.

4. In the absence of (1) (2) and (3), the email address for the records department is collected.

5. If none of the above email addresses are publicly available, any email address available on the the police department website is collected.

6. If there are no email addresses available on the department website, a cursory search is performed to find email contacts on other related websites (e.g., the website of the city that a department is located in or the official Facebook page for a department)

[25] In other instances, there were no publicly available email addresses associated with a police department of interest.

B Online Appendix: Identity Construction

Six different "identities" will be used:

1. Black Female
2. Black Male
3. White Female
4. White Male
5. Hispanic Female
6. Hispanic Male

Consistent with standard practices in the correspondence study literature, identity (gender and race/ethnicity) of the email sender will be implied by name (first name and last name). Ten unique first names and six unique last names are chosen for each identity (60 unique name combinations for each identity). Using multiple names for each identity minimizes the importance of a specific name.

- First names are selected from research done by Gaddis (Gaddis (2017a), Gaddis (2017b)). The top ten most racially identifiable first names (when coupled with last names), are chosen.
- Last names are selected from the 2010 Census. Three criteria are used to select last names:
    1. Percent of persons with that name having a specific race/ethnicity (e.g., White)
    2. Percent of persons with that name having the other relevant race/ethnicity (e.g., Black or Hispanic)
    3. The rank of the name (i.e. how common the last name is in the United States)

Name Search Equation: I selected surnames for this experiment that were both (1) racially distinctive and (2) commonly found. Priority was given to racially distinctive, because of the importance of race in the design of the experiment. However, I also wanted to avoid the scenario where police departments act differently if they see an exceedingly uncommon last name. In other words, I want race, and only race, to be communicated by the name of the identity. The three equations below reflect the priorities I used to select the names. I decided it was unnecessary to difference the Hispanic surnames with the other two groups because of how uncommon it was for Black and White people to have a surname commonly used by Hispanic people.

- For Black Names: $percent\ race_{black} - percent\ race_{white}) - .05 \times rank_{black\ name}$
- For White Names: $percent\ race_{white} - percent\ race_{black}) - .05 \times rankwhite\ name$
- For Hispanic Names: $percent\ race_{hispanic} - .05 \times rankwhite\ name$

The full list of names can be inferred by the following two tables (there are 360 unique name combinations). Six high-profile recognizable celebrity names were omitted: Denzel Washington, Tyra Banks, DaShawn Jackson, Seth Meyer(s), Katelyn Olson and Pedro Martinez. These names have widespread recognition and during the testing process, respondents noted that they strongly associate these names with the celebrities having the same name.

Last Names		
White	Black	Hispanic
Olson	Washington	Hernandez
Schmidt	Jefferson	Gonzalez
Meyer	Jackson	Rodriguez
Snyder	Joseph	Ramirez
Hansen	Williams	Martinez
Larson	Banks	Lopez

First Names					
White Male	White Female	Black Male	Black Female	Hispanic Male	Hispanic Female
Hunter	Katelyn	DaShawn	Tanisha	Alejandro	Mariana
Jake	Claire	Tremayne	Lakisha	Pedro	Guadalupe
Seth	Laurie	Jamal	Janae	Santiago	Isabella
Zachary	Stephanie	DaQuan	Tamika	Luis	Esmeralda
Todd	Abigail	DeAndre	Latoya	Esteban	Jimena
Matthew	Megan	Tyrone	Tyra	Pablo	Alejandra
Logan	Kristen	Keyshawn	Ebony	Rodrigo	Valeria
Ryan	Emily	Denzel	Denisha	Felipe	Lucia
Dustin	Sarah	Latrell	Taniya	Juan	Florencia
Brett	Molly	Jayvon	Heaven	Fernando	Juanita

As mentioned, the first names were selected from Gaddis (2017a) and Gaddis (2017b). In these studies, Gaddis analyzes the correlation between the average level of the mother's education for a given first name and accuracy of perceived race and ethnicity of that name. For instance, Black names associated with lower education levels for mothers are more often perceived as Black than Black names associated with mothers with higher average education levels. In my study, while creating the identities, the associated maternal education levels documented by Gaddis are recorded in my database.

C Online Appendix: Email Account Creation

To implement this study, sender email addresses had to be created for each putative identity. Ideally, each of the 360 identities would have a unique email address. During the pre-testing process, respondents suggested that
"firstname.lastname.birthyear@mail.com" was the most realistic email address template. However, due to constraints from popular email servers (e.g., Yahoo), this was not feasible. Instead, a unique account was made for each *last name* (18 accounts in total). Due to availability, I had to be creative in creation of the email address. All of the addresses include some version of the relevant last name.[26] Often included is a birth year (e.g., **Banksss.1991@mail.com**). I do not expect that the implied birth year will be a salient component of the email, but I will make a cursory examination of the role that the email sender's apparent age plays in response rates for police departments.

[26]Due to the prevalence of people with the last names chosen for the study, it was often difficult to find available addresses with the specific last name. As a result, I had to make creative decisions to create a plausible and name-relevant address. For example, "h3rnandez.1973@mail.com".

D Online Appendix: Example Email

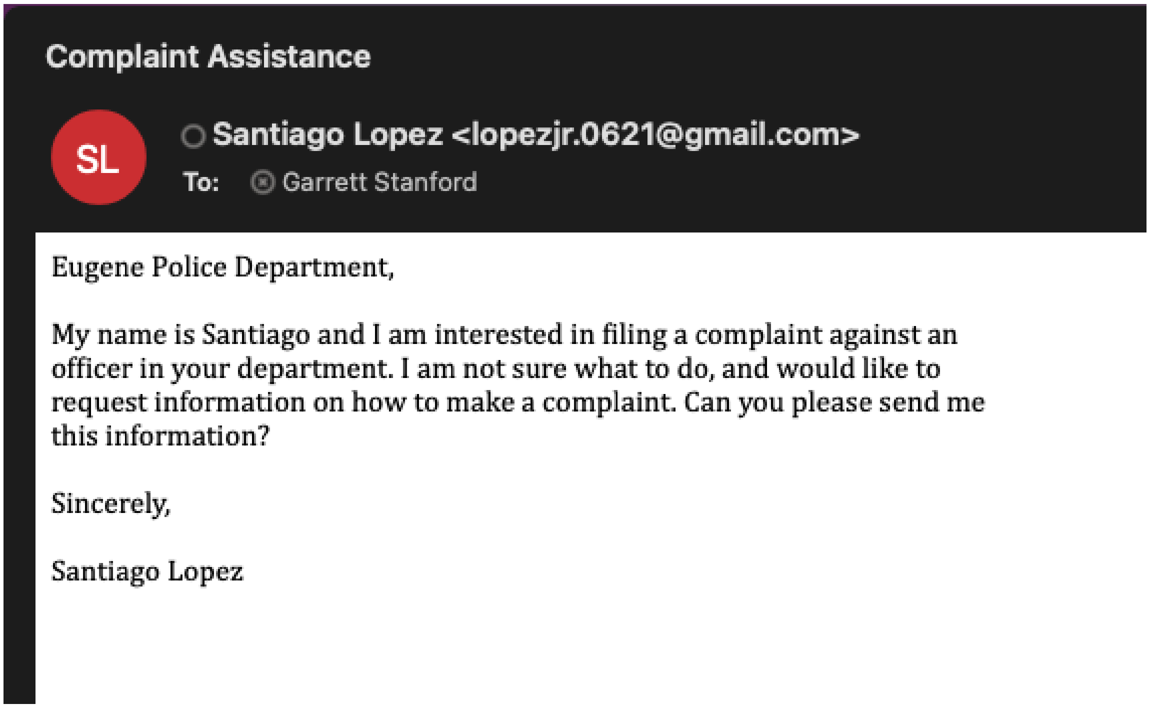
D.1 Email Text

The body of text for the email has been developed in consultation with other economists and a legal expert. The primary criterion in creating the right text for these emails concerned plausibility—i.e., I needed to create an email that sounded like a genuine request from a real citizen. Drafts of the email were sent to colleagues and police departments not selected for the correspondence study to assess the plausibility of the email. The body of the email message template reads as follows:

> ***Police Department Name**,*
>
> *My name is* ***first name*** *and I am interested in filing a complaint against an officer in your department. I am not sure what to do, and would like to request information on how to make a complaint. Can you please send me this information?*
>
> *sign off*
>
> ***full name***

Where *full name* includes a first and last name, and *sign off* is randomly assigned as "Thank you!" or "Sincerely,". The decision to exclude a "Hi" or "Hello" was based on the increased likelihood of the email being filtered as spam during the preliminary testing process mentioned above. [27]

[27] There is a small concern about this email being rejected as implausible. For example, a very small police department might know everyone with whom they have recently interacted and would be able to deduce, with little effort, that the email is fabricated. A small police department might also be more likely not to respond to an email because of staffing limitations. However, because assignment of treatment (see below) is balanced across departments, estimates should be remain unbiased. In future research, an alternative email to departments with a more innocuous inquiry (e.g., "Do you have a lost-and-found?") could shed light on the matter.



Complaint Assistance

SL

Santiago Lopez <lopezjr.0621@gmail.com>

To: Garrett Stanford

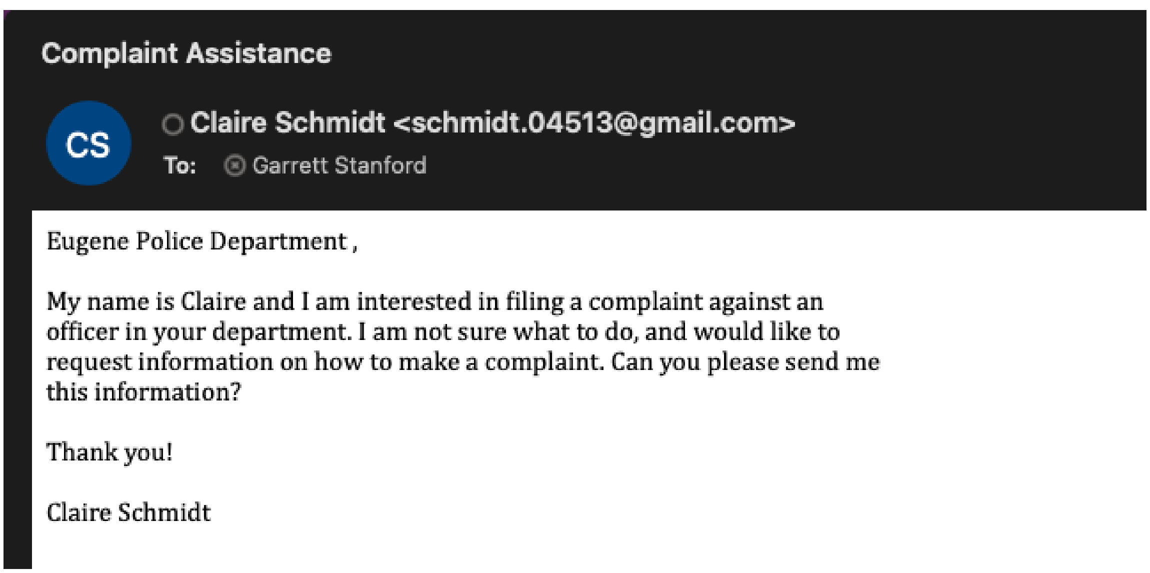
Eugene Police Department,

My name is Santiago and I am interested in filing a complaint against an officer in your department. I am not sure what to do, and would like to request information on how to make a complaint. Can you please send me this information?

Sincerely,

Santiago Lopez

Figure B1: Example email



Complaint Assistance

CS

Claire Schmidt <schmidt.04513@gmail.com>

To: Garrett Stanford

Eugene Police Department ,

My name is Claire and I am interested in filing a complaint against an officer in your department. I am not sure what to do, and would like to request information on how to make a complaint. Can you please send me this information?

Thank you!

Claire Schmidt

Figure B2: Example email

E Online Appendix: Treatment Assignment

Police departments are randomly assigned the sender identity they will see. The first step of treatment assignment was to balance the number of departments by state each week, so that every state received roughly the same number of emails each week. Next race and gender treatment are randomly assigned within state, with race and gender treatment levels balanced within each state. Given that assignment of emails to department by week within state was randomized, race and gender assignments are independent of week. Additionally, race and gender are roughly balanced across weeks—also as a result of the randomization of all treatment components. After week, gender and race are assigned, day of week is randomly assigned. Next, the sign off for each email is randomly assigned (the email sign off can be either "Thank you" or "Sincerely" followed by the sender's name). The actual assignment of email sender first and last names to each department is randomized across all weeks and states.

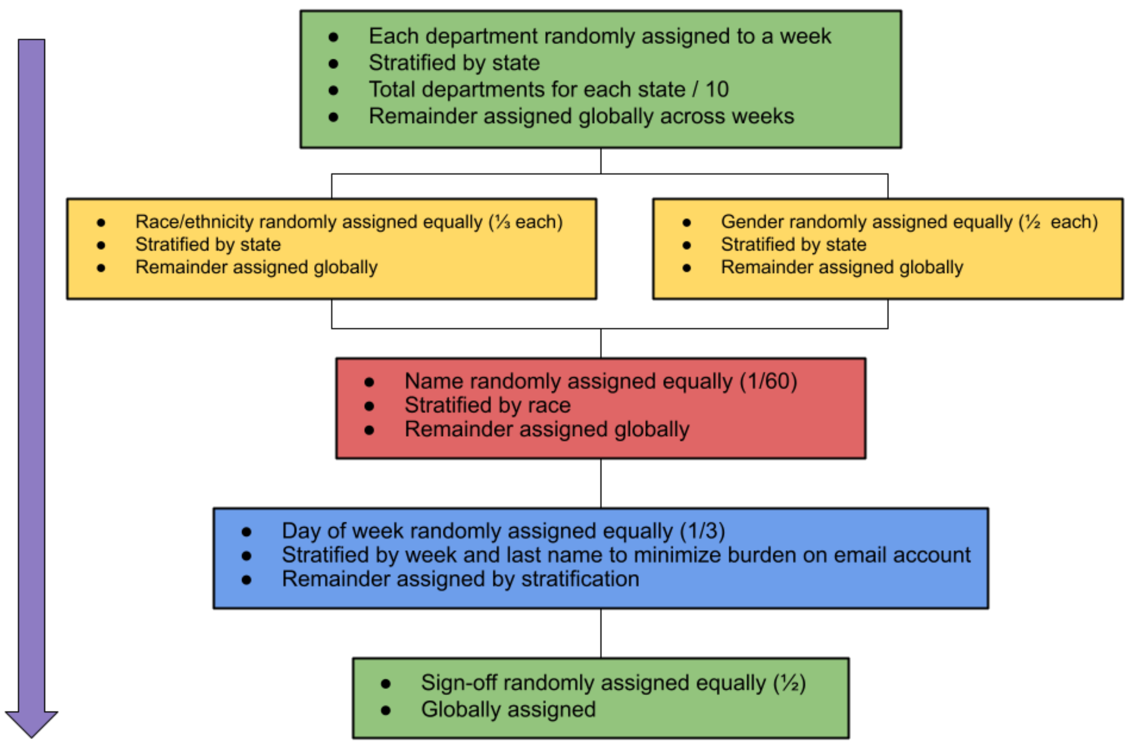


Table B1: Distribution of Race, Ethnicity and Gender identity assignment by state

State	Putative Identities					Response Statistics	
State	Black	White	Hispanic	Male	Female	Total	Mean
AK	3	2	2	4	3	5	0.714
AL	5	5	5	8	7	10	0.667
AR	6	4	6	7	9	8	0.500
AZ	6	5	4	9	6	12	0.800
CA	26	22	19	38	29	46	0.687
CO	7	7	5	10	9	14	0.737
CT	11	9	16	20	16	21	0.583
DE	2	1	1	1	3	2	0.500
FL	17	19	18	30	24	35	0.648
GA	9	10	6	10	15	14	0.560
IA	6	6	4	5	11	6	0.375
ID	3	3	5	6	5	6	0.545
IL	28	26	29	39	44	61	0.735
IN	10	10	11	16	15	19	0.613
KS	5	7	6	11	7	11	0.611
KY	3	5	4	7	5	6	0.500
LA	2	6	4	7	5	6	0.500
MA	16	17	19	28	24	31	0.596
MD	7	5	2	8	6	7	0.500
ME	6	5	6	8	9	10	0.588
MI	14	18	12	22	22	27	0.614
MN	12	8	11	13	18	24	0.774
MO	9	9	10	16	12	16	0.571
MS	6	5	5	7	9	4	0.250
MT	2	1	3	1	5	3	0.500
NC	10	6	11	12	15	17	0.630
ND	3	1	1	2	3	4	0.800
NE	4	6	2	5	7	10	0.833
NH	4	4	6	7	7	9	0.643
NJ	27	30	25	41	41	45	0.549
NM	5	3	4	7	5	3	0.250
NV	2	1	2	3	2	2	0.400
NY	14	18	19	26	25	31	0.608
OH	25	21	32	37	41	49	0.628
OK	6	7	3	8	8	10	0.625
OR	7	6	10	10	13	17	0.739
PA	25	22	21	33	35	44	0.647
RI	3	3	4	4	6	5	0.500
SC	5	6	4	10	5	7	0.467
SD	1	3	1	3	2	4	0.800
TN	6	6	6	10	8	12	0.667
TX	31	20	31	41	41	60	0.732
UT	6	4	2	8	4	9	0.750
VA	6	3	3	6	6	8	0.667

Continued on next page

Table B1 – continued from previous page

VT	2	3	3	3	5	6	0.750
WA	11	9	9	13	16	20	0.690
WI	10	10	11	17	14	26	0.839
WV	3	4	4	7	4	4	0.364
WY	1	2	3	4	2	4	0.667

Table B2: Distribution of Race, Ethnicity and Gender identity assignment by week

Week	Putative Identities					Response Statistics	
week	Black	White	Hispanic	Male	Female	Total	Mean
1	75	73	68	107	109	139	0.644
2	74	79	63	103	113	141	0.653
3	59	68	77	103	101	132	0.647
4	74	60	76	119	91	123	0.586
5	77	62	78	104	113	144	0.664
6	79	71	68	112	106	131	0.601

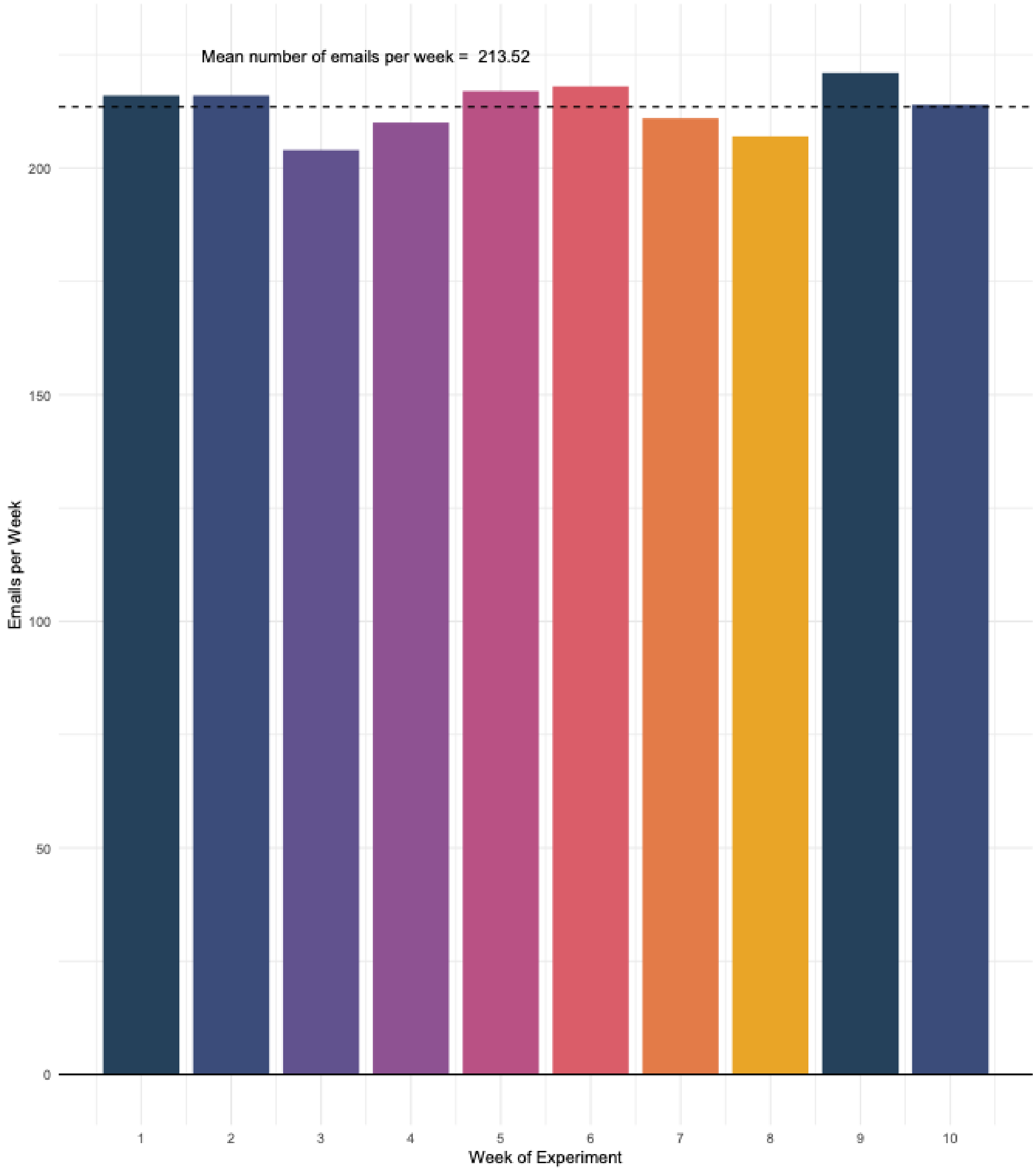


Figure B1: Emails sent by week.

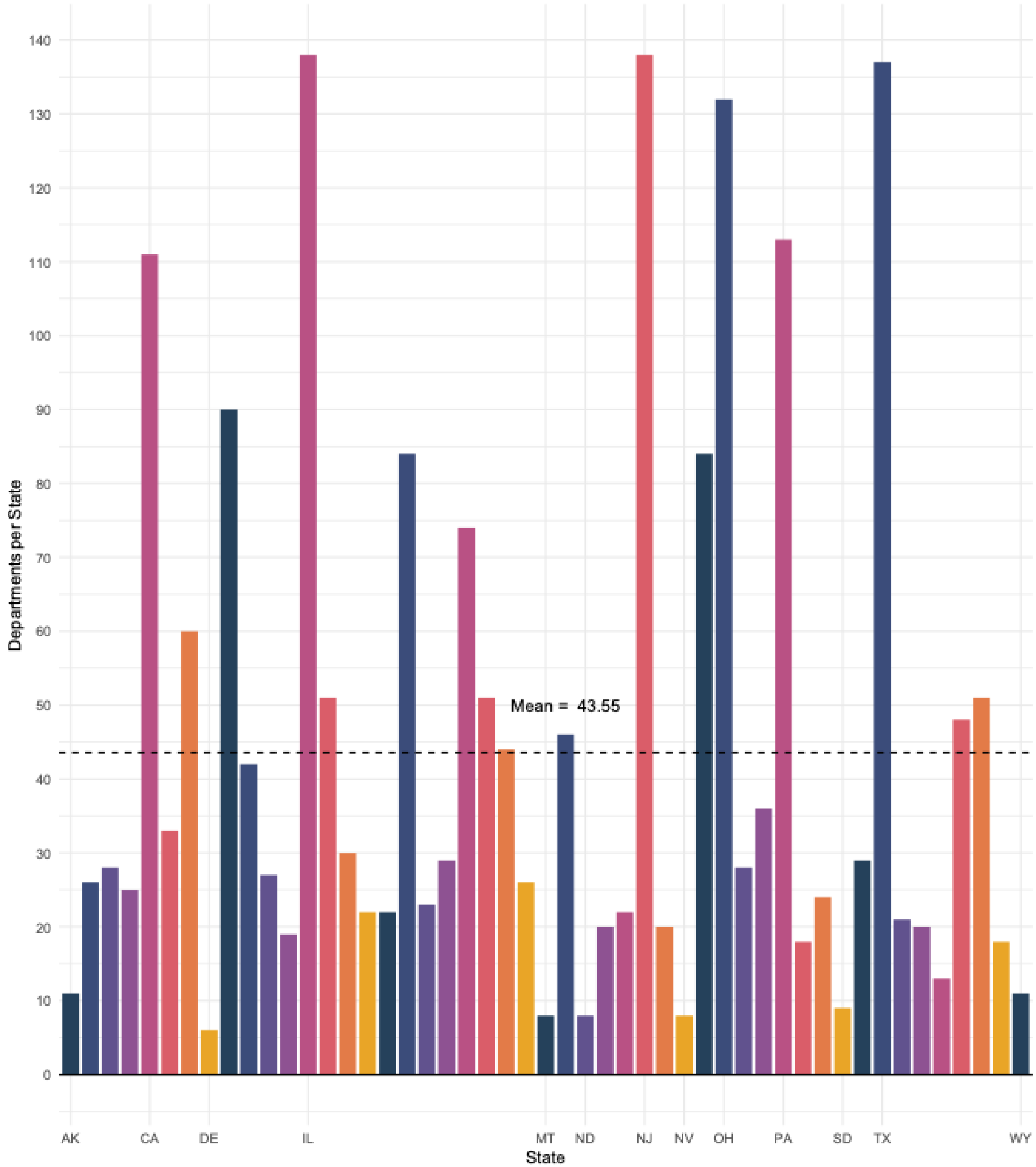


Figure B2: Departments by state included in the study.

F Online Appendix: Experiment Implementation

I created eighteen email accounts—one for each last name. The accounts were then linked to Mozilla's Thunderbird mail application to help automate the emailing process.[28] In Thunderbird, for each email address, 20 identities were created (10 females and 10 males). Although the email address that is seen by police departments cannot be arbitrarily manipulated, the "name" of the sender can be changed from message to message. For instance, an email can be sent as **Claire Olson <olson2292@mail.com>** or **Hunter Olson <olson2292@mail.com>**. This helps increase the salience of the putative identity and decrease attention to the less-specifc email address itself.

Each department will receive just one email. Emails will be sent over a ten-week period. Spreading out the randomized controlled trial (RCT) over 10 weeks insures against the possibility that unique unanticipated current events could plausibly affect police department behavior (e.g.,, a high-profile regional or national incident involving the police). In the case of a high-profile policing incident, a weekly roll-out of the emails will allow me to detect the possible effect of any such event on police departments' responses to the emails.

The timing of the roll-out is randomly selected using the following procedure. Police departments are randomly assigned to one of the ten weeks, while being stratified proportional to the total number of departments in each state. Each state's total police departments (in my data set) are split into 10 equal groups and assigned to a week. In the event that, after the initial assignment, the number of departments by state are not divisible by 10, the remainder of the police departments are randomly assigned across the weeks. In the event that the total number of departments from a state is less than 10, departments are randomly assigned to the ten different weeks (with a maximum of one department per week). Each putative sender identity (i.e. email address) has the same probability of being assigned to any one of the 10 weeks.

During each week, the emails are sent on Monday, Tuesday and Wednesday. Assignment of weekday is randomized. The decision to choose different days is largely motivated by an effort to improve the ease of implementation of the emailing process for the researcher. Each email must be sent individually, so

[28] I had originally intended to use the mailR package from R, but due to increased security policies with many popular email servers, that option is no longer as user friendly. To use mailR with, for example, Google, one needs to change the Google account settings to allow "less secure apps". However, as of May 31st, this setting can no longer be adjusted. There are possible workarounds, but I decided to adopt an alternate strategy.

it proved easier for me to monitor the email process by spreading out the emails over a few days (with roughly 70 emails being sent each day).

All emails are sent at roughly 9 a.m. local time according to the time zone of the police department in question. However, if for a given week and given day, the same email sender address is being used for more than one police department (as dictated by the random assignment of race), a five-minute delay between each email from the same address, independent of first name, is employed. The strategy is adopted so that a single putative email account does not have to send more than one email at an exact time (i.e. at exactly 9 a.m.).

G Online Appendix: Response Time and Word Count

Not all responses are created equally: The current analysis of the data from this correspondence study designates the outcome variable to be a police department's timely non-automated response to a request for help. Consequently, the results are a coarse reflection of the average department's willingness to respond to a citizen's request for help in making a complaint about an officer. However, the premise of biased policing refers to both the frequency of interaction between officers and citizens, as well as the conduct during the interaction. Even in the specific context of an email request for a complaint form, detecting and understanding potential differences in department behavior across different sender identities is worth exploring. For example, conditional on a department providing any response, do responses differ in their helpfulness and tone across identities and, if so, how do they differ? In some instances, scrutiny of verbatim department responses reveals that not all departments are willing to guide the citizen to the officer-complaint forms. In other instances, departments specifically advise against making a formal complaint. Responses also tend to reflect a wide range of sentiment. Some departments include an apology on behalf of the department, while others simply send a phone number with no other information—the assumed implication being that the complainant should call that number for assistance. To begin to answer the question of differential response conditional response, a cursory examination of heterogeneity of responses is performed.

Table B1: Response time and word count of response measured across identities

Dependent Variables:	Word Count	Response Time (hours)
Model:	(1)	(2)
Variables		
White × Female	0.2948	-1.653
	(3.844)	(2.573)
Hispanic × Male	3.909	3.032
	(2.533)	(3.103)
Hispanic × Female	-3.051	-4.157
	(3.937)	(3.056)
Black × Male	4.518	8.778
	(7.728)	(7.215)
Black × Female	-4.038	2.567

Continued on next page

Table B1 – continued from previous page

	(2.849)	(3.919)
Fixed-effects		
Week	Yes	Yes
State	Yes	Yes
Fit statistics		
Observations	1,413	1,413
R^2	0.08298	0.04623
Within R^2	0.00291	0.00892

Clustered (week & dept_address_state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table B1 reports the differentials of (1) the word count of emails from the departments and (2) the time it takes for a department respond between White male identities and the other five identities. Table B1 suggests that conditional on response, at least on the two specified dimensions, there does not seem to be any evidence of discrimination. There are a few reasons to not make strong conclusions about these null results. Most importantly, the analysis is subject to selection bias. These results are based only on the departments that *do* respond, which are different than the departments that do not respond. Additionally, word count is a crude measure of helpfulness and sentiment. An email could be helpful, friendly and to the point, but still would reflect a word count similar to an email that is unhelpful and/or unfriendly. Time of response is a stronger indicator of helpfulness. However, a quick response could be the result of a department eager to help or a department being reactive to an accusation against one of their officers. To get a strong understanding of differences in the helpfulness and sentiment of responses would require selection correction and a more rigorous sentiment analysis.

H Online Appendix: Summary Statistics

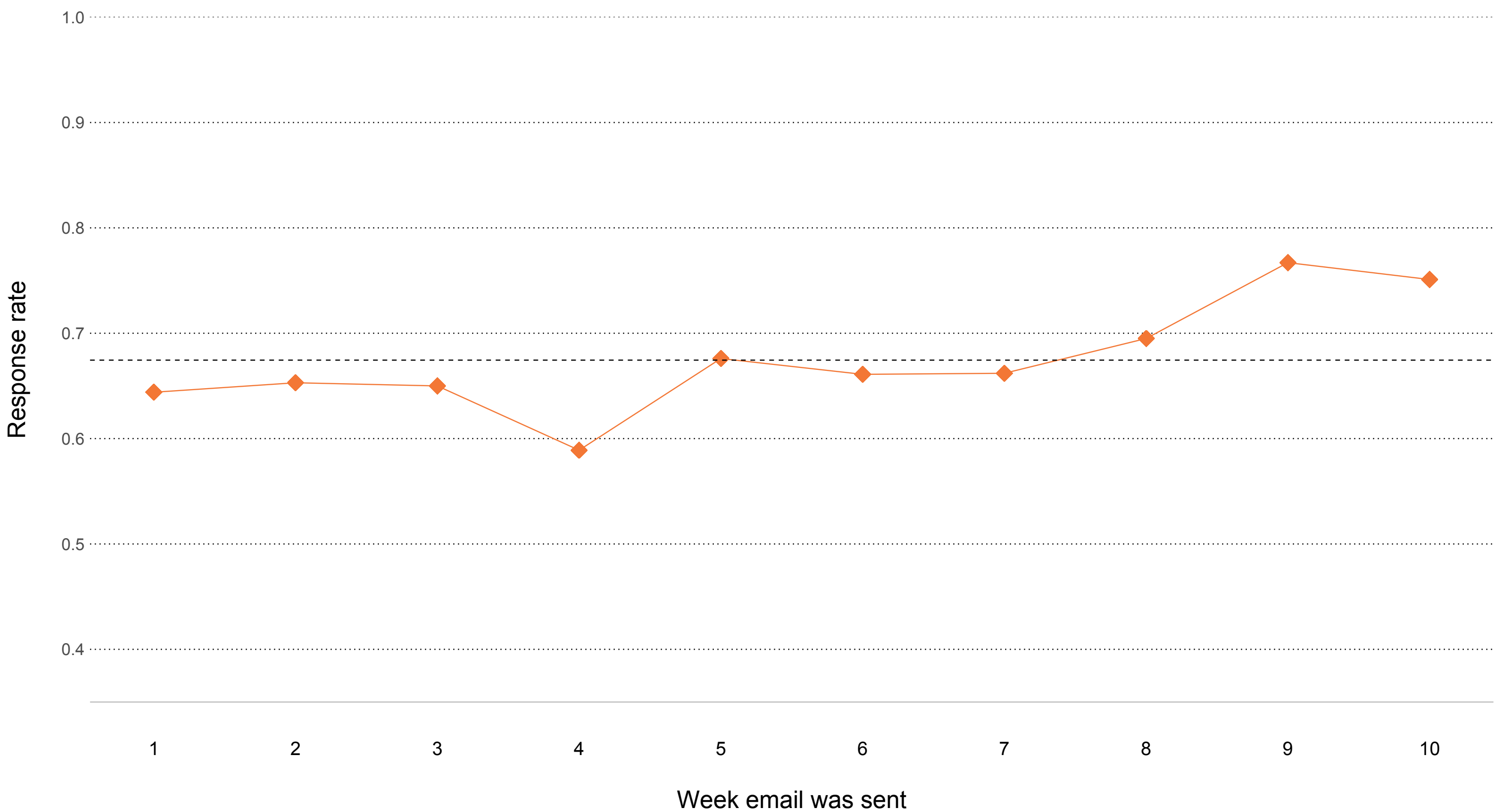


Figure B1: The mean response rate by week by identities. Mean response rate across all weeks and identities (66%) is depicted by dotted black line.

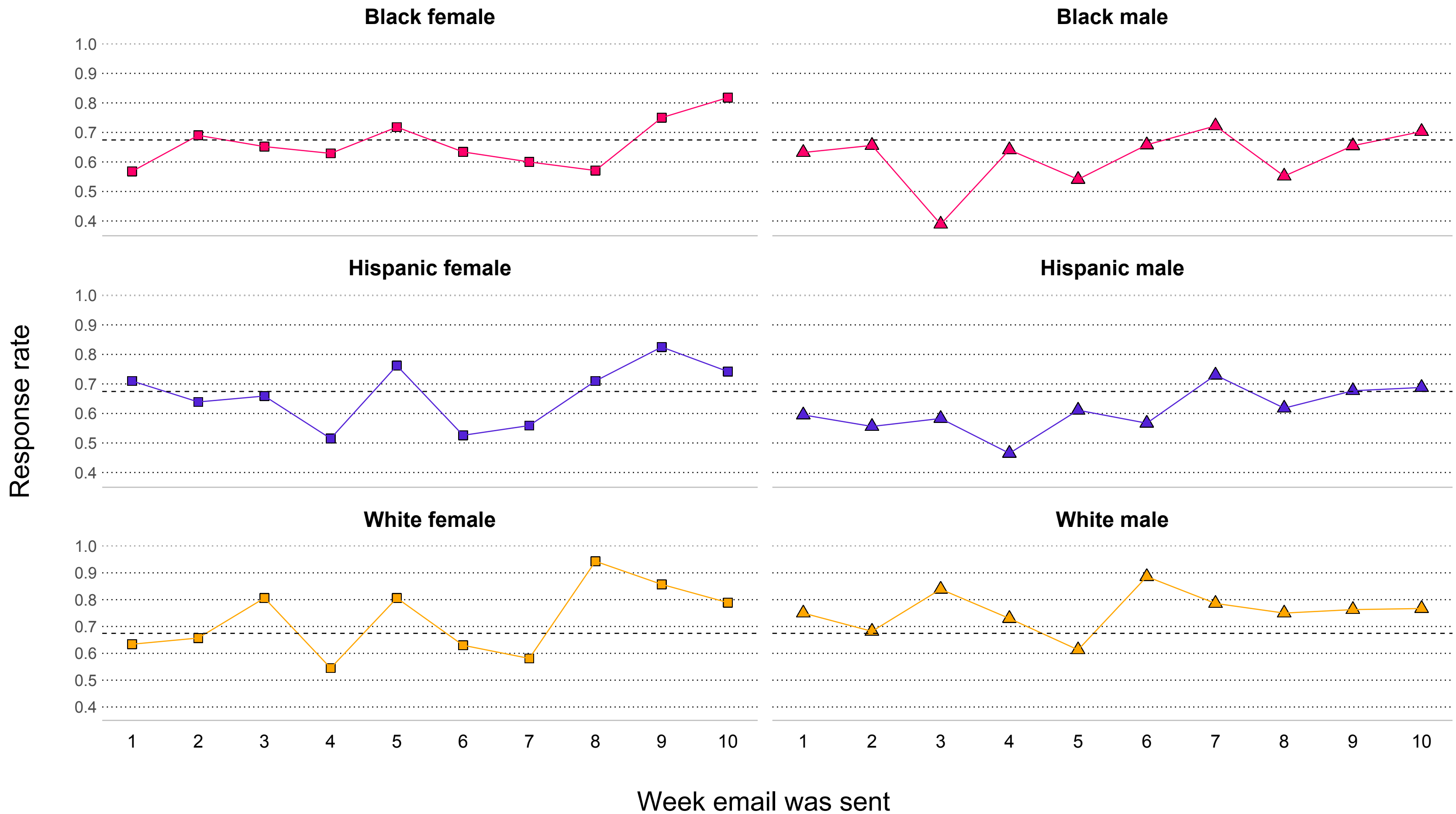


Figure B2: The mean response rate by week for all putative identities

I Online Appendix: Department Size

J Department Size

Comparing Columns 1 and 2 in Table 6 it is apparent that weighting by the population of the police department's jurisdiction exacerbates the differences in response rates (with the exception of Hispanic male response rates). One possible explanation for these results is that departments serving larger populations are more likely to discriminate. To test this interpretation, departments are separated into five bins determined by the quintiles of the local populations of the departments included in the study. Model 1 from 5 is re-estimated, but this time interacting both *Hispanic* and *Black* with the five population bins. Figure B1 shows the results of this exercise. Figure B1 reveals no clear pattern in the relationship between population size and response rate. Black identity response rates are lowest for the largest and smallest quintiles, but the pattern does not hold for Hispanic identity response rates.

The relationship between police department local population and response rate is ambiguous. Notwithstanding a clear relationship between local population and propensity to discriminate, heterogeneous response rates across population sizes suggest that studies that restrict their area of focus to a limited number of local governments, or at least similarly sized populations, may not be able to extend their results to smaller (or larger) populations. Furthermore, the higher differentials in response rates between White and non-White response rates when weighting by population (Table 6) indicate, in a crude measure, that more people are discriminated against than is evident by finding the level of discrimination for the average police department.[29]

A variable strongly correlated with local population size is the number of employees working for a police department. Mechanically, as local populations increase, so do the sizes of departments. For the departments include in the present study there are on average 3 employees for every 1,000 residents. Results from Table 7 suggest that bigger departments are more likely to discriminate against Black identities and White female identities, and discriminate less against Hispanic identities. When the results are weighted by

[29]The design of this study only sends one email to each department, so a department responding to a White email does not necessarily mean the same department would *not* respond to a non-White email. Conversely, a department not responding to a non-White email does not necessarily mean the same department *would* respond to a White email. However, higher levels of discrimination for departments serving larger populations do suggest that a larger share of the population might be discriminated against than is indicated by the unweighted results.

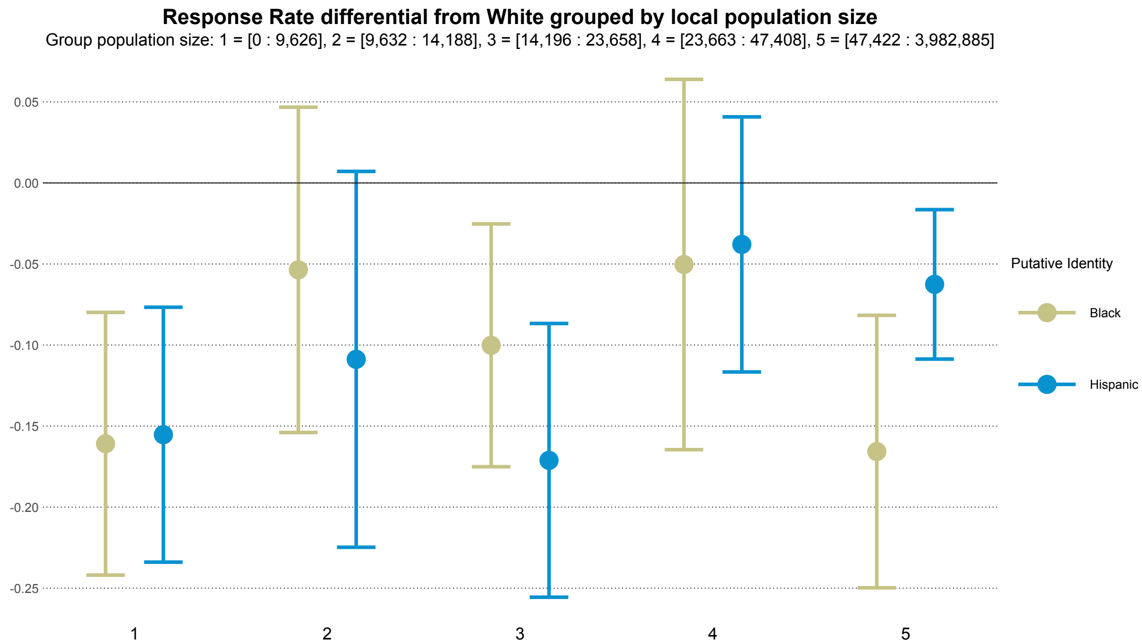


Figure B1: Response rates differentials for Black and Hispanic identities from White identities separated into bins determined by local population size.

population in Columns 3 and 4 in 7, the discrepancies grow, which makes sense given that agency size is correlated with population size.[30] To disentangle the effect of agency size and population size on response rate given their the two sizes correlations, the same exercise from Figure B1 is done. This time the quintiles are determined by the number of employees divided by local population size—a per capita measurement. Figure B2 displays the results of this exercise.

No clear pattern emerges for Hispanic or Black identity response rates. Compared to Figure B1, Figure B2 has smaller differences for the first bin. However, in both Figures the biggest bin, bin 5, exhibits large differences for Hispanic and Black response rates. It is conceivable that the low response rates for departments with big populations might be attributed to departments being overextended and thus less capable of responding to requests for assistance. However, the highest employee per capita departments exhibiting the highest degree of discrimination runs counter to that argument. It is concerning that departments most capable, in terms of employees per capita, to respond to requests are most likely to discriminate.

[30]One curious result of Table 7 is comparing the point estimate for Hispanic females from Column 2 and Column 4.

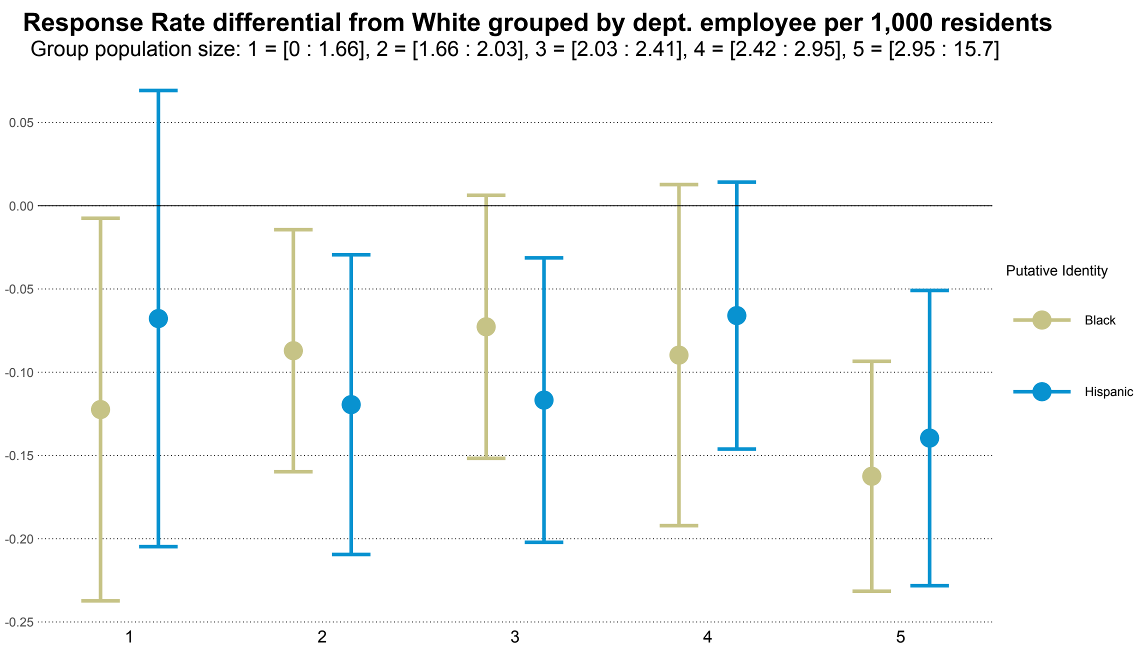


Figure B2: Response rates differentials for Black and Hispanic identities from White identities separated into bins determined by local population size.