"I'd like to make a complaint against an officer": Field experiment evidence of police discrimination based on citizen race, ethnicity, and gender
Garrett Olson Stanford
October 2022
JEL No. C93, J15, J16, K40

**ABSTRACT**

Despite considerable national attention, research remains inconclusive about the existence of biased policing. I use an RCT to statistically test whether police departments respond differently to citizen requests for information based on the perceived race or gender of the requester. In the form of a correspondence study, I send email requests to more than 2,000 U.S. police departments asking for information about how to lodge a formal complaint about an officer in each department. I systematically vary the putative race/ethnicity (Black, Hispanic and White) and gender (male and female) of the email sender. I compare response rates for the putative identities as evidence of biased policing. I then examine which, if any, observable departmental or jurisdictional characteristics (e.g. size of agency and local population) are related systematically to response behavior. One of the most valuable features of the the study's design is that I am able to assess departments for evidence of biased policing without relying on administrative police department data. Results indicate that police are significantly less likely to respond to requests from Black and Hispanic putative identities than to requests from White putative identities. Response rates are highest for White male identities and lowest for Black male and Hispanic male identities.

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

# 1 Overview

## 1.1 Criticisms of Police

There is a strong consensus that law enforcement agencies are an integral part of American society, and research has shown that police can be effective at reducing crime (e.g., Chalfin et al. (2021), Chalfin and McCrary (2018), Chalfin and McCrary (2017), Cheng and Long (2018), Evans and Owens (2007), Mello (2019), Weisburst (2019b), Weisburd (2021)). However, there is a longstanding and heated debate concerning the ultimate impact that some police practices can have on social welfare. Of particular concern is the presence of bias in police behavior—especially racially motivated bias. In 2014, outrage over racially-biased policing boiled over following the death of Michael Brown in Ferguson, Missouri[1]. Six years later, in 2020, protests broke out across the nation after police killed George Floyd. Despite the durable presence of police reform in the national dialogue and despite alarmingly frequent, but anecdotal, incidents in the news, most studies that attempt to document incidences of biased policing lack the statistical rigour required to infer causal relationships (Smith et al. (2017)).

## 1.2 Summary of study

This paper presents the results of a randomized controlled study in which more than 2,000 police departments were emailed and asked for help concerning how to make a complaint about an officer in that department. Each police department was emailed by a fictitious citizen. The email that every department received was identical except for a randomly assigned putative race/ethnicity (Black, Hispanic and White) and gender (male and female) for the fictitious email sender. Race/ethnicity and gender were communicated through the first and last name of the putative email sender. The response rates for these emails were then compared for the different identities to determine if police departments were systematically more likely to respond to a citizen of a particular race/ethnicity or gender than to another.

[1] US Department of Justice Civil Rights Division (2015)

## 1.3 Evidence of Bias

A growing body of research has highlighted the disproportionate burden that policing can place on non-White citizens. This research covers a variety of contexts: general arrest rates (e.g., Bulman (2019)); predatory fines and asset forfeitures (e.g., Makowsky et al. (2019), Sances and You (2017), Shoub et al. (2021), Singla et al. (2020), Su (2020), Su (2021)) ; police stops and searches (e.g.,Antonovics and Knight (2009), Bandes et al. (2019), Feigenberg and Miller (2021), Gelman et al. (2007), Goel et al. (2016), Pierson et al. (2020)); traffic enforcement (e.g.,Goncalves and Mello (2021) and West (2018)); and use of force (e.g.,Edwards et al. (2019), Fryer (2020), Nix et al. (2017), Ross (2015)). The extent to which racially biased policing can unfairly impact individuals and sometimes entire communities can be quite significant. For instance, Edwards et al. (2019) find that police use of force is one the leading causes of death for young men of color in the United States. Most studies that address biased policing, while strongly suggesting the existence of bias, are unable to identify causality. Thus researchers remain divided on the existence of biased policing (Smith et al. (2017), and Fridell (2017)). Furthermore, although some researchers have managed to use research designs to permit causal inferences, a significant caveat to the vast majority of these studies is that the data are often supplied by the police themselves. Reliance on police-reported data can lead to inconclusive or incorrect conclusions if police departments strategically or unintentionally misreport (e.g., Knox et al. (2020) and Luh (2019)).

## 1.4 Motivation for correspondence study

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 (Gaddis (2017a), Bertrand and Duflo (2017)). In a correspondence study, individuals (often fictitious) that are identical for all observable characteristics except the characteristic of interest apply for a service or good. The researcher then examines whether the varied characteristic has an effect on the outcome of the application. [2] Despite the

[2] A correspondence study refers to a certain type of audit study. Bertrand and Duflo (2017) define audit studies and correspondence studies:

> Audit studies send out individuals who are matched in all observable characteristics except for the one in question (race, criminal record, etc.) to apply for jobs or make purchases, then researchers analyze the responses. Correspondence

increasing popularity of correspondence studies to study discrimination, and given the dearth of causally identified bias in policing, this paper is still one of the first studies to use a correspondence study to identify the presence of racially and gender motivated bias in policing.[3]

## 1.5 Importance of this study

By using a correspondence study, this paper overcomes two of the main challenges faced by other researchers seeking to study discrimination in the context of law enforcement: (1) the need for causal identification, rather than just correlation and (2) potentially compromised reporting in data collected and/or provided by law enforcement agencies.

As a type of RCT, the results of a correspondence study, (if it has sufficient statistical power and treatment is sufficiently randomized), can be reasonably assumed to be causal. In the context of police discrimination, a serious obstacle for researchers is the need to disentangle how much of police activity that disproportionately targets non-Whites stems from biased policing and how much is explained by different levels of participation in criminal activities (Fridell (2017)). The present study avoids this challenge by creating a citizen-initiated police interaction that is not predicated on a crime taking place.

Avoiding the use of data provided by police departments has a number of significant advantages. Police data can be unreliable for a number of possible reasons. First, departments, either nefariously or inadvertently, can have ongoing difficulties in accurately reporting data. Second, in a similar vein, police data can be a product of subjective reporting by individual police officers and department-specific classification standards. Even when police officers accurately and honestly record officer conduct, decisions made in the heat of the moment during a citizen-officer interaction can influence how events are recorded. Finally, departments may be unwilling to disclose what might be deemed sensitive information.[4]

This study also touches on accountability, another important issue that departments frequently come un-

studies—which represent by far the largest share of field experiments on discrimination—do the same but control for more variables by creating fictitious applicants (often for jobs or apartments) who correspond via mail.

In the case of this experiment, email is used instead of the traditional approach of "snail mail," and—as explained below—a request is made for assistance from police departments, as opposed to applying for jobs or making purchases.

[3]The only other correspondence study examining racial bias in law enforcement is Giulietti et al. (2019)—discussed below

[4]Weisburst (2019a) do not find evidence of racially biased policing in Dallas, but 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—or at least the department's data reflects as much in Weisburst (2019a)

der criticism for. By asking departments for help in making a complaint about an officer in their department, the present study examines the intersection between biased policing *and* the willingness of departments to hold their officers accountable—an essential concern for policy makers interested in reforming law enforcement in the United States. Additionally, this study appears to be the first RCT to study discrimination in the context of law enforcement for Hispanic citizens and for discrimination across genders (male and female).[5]

# 2 Experimental Design and Data

## 2.1 Experiment

### 2.1.1 Department Selection

The police departments included in this study are a stratified random sample from all United States departments. To accomplish this, I randomly sampled the US Census's universe of local governments in batches of 1,000.[6] From the sample of local governments, I searched the internet for an email address for the corresponding police department. Some local governments did not have their own police departments, and some police departments did not have publicly available email addresses. When a department had multiple email addresses publicly available, priority was given first to the general department email, then to the police chief and then to the next highest in command officer. In the end, 2,135 departments were selected to receive emails.

### 2.1.2 Identity Creation

Six putative identities were created for this study: Black female, Black male, Hispanic female, White female and White male. Each identity consisted of 60 unique first-last name combinations. Last names were selected from the US Census. I selected last names for this study that were both racial distinctive and commonly found. Six last names were selected for each race. First names for each putative identity were selected from Gaddis (2017b) and Gaddis (2017a). I chose the ten most racially distinctive names for the

[5] Weitzer (2014) points out that considering the growing population of Hispanic Americans, the lack of research on police-citizen for Hispanics is "particularly puzzling."

[6] The universe of local governments were filtered to exclude counties and governments with populations less than 7,500.

respective identities. In total, 360 unique names were created for this study (6 identities $\times$ 6 last names $\times$ 10 first names).[7] After the names for each identity were selected, I created a unique email address for each last name used in the study. The email address included the identities first name.[8]

### 2.1.3 Email

Each department received one email from a randomly assigned putative identity. All emails are identical with two exceptions (1) the name of the email sender and (2) the sign-off used in the email. [9] The name of the email sender is used twice in each email to increase the salience of the name.[10]

### 2.1.4 Timing

This study was conducted over a ten-week period, from late June 2022 to late August 2022. Each week roughly 200 emails were sent out, split across Monday, Tuesday and Wednesday. All emails were sent at approximately 9 a.m. local time for each police department. A ten-week sampling period was selected to minimize the chance that a single adverse event—in the national or regional news— would compromise the generalizability of the results. Splitting the emails across days of the week was done to reduce the difficulty of sending the emails.[11]

### 2.1.5 Treatment Assignment

Treatment for this study is the putative identity (race and gender) each department sees. Treatment was first stratified by week and by the state the departments are located in, so that the number of departments by state are balanced during each week. Treatment was then randomly assigned across departments within each week-state strata. For more details on the treatment assignment process and results of randomization see Appendix D.

[7] Please refer to Appendix A for details on the names used in this study.

[8] Please refer to Appendix B for details on the email addresses used.

[9] The sign-off was randomly assigned sign-off across emails. The sign-off varied between cordial sentiment (*Thank you!*) and curt sentiment (*Sincerely,*).

[10] Please refer to Appendix C for examples of the email text.

[11] Please see Appendix F for more details on how the experiment was implemented.

## 2.2 Data

Several datasets were used to accomplish this study. As mentioned, data from the US Census was used to create a pool of local governments in the department selection process. Local governments eligible for inclusion in the study were excluded state and county governments and were limited governments with populations of at least 7,500 residents.[12] Departments selected for inclusion in the study were then matched with police department directories from OpenPolice.org and ICPSR. Department directories added information about agency location and unique identification numbers.

Data for several other observable characteristics were included in the study that ex ante seemed potentially important determinants in the response behaviors of police departments. The number of officers and civilian employees for each department; county-level income information; and county-level race/ethnicity makeup. Employee counts for each department was provided by UCR data codified by Jacob Kaplan. The UCR data includes employee counts through 2020. However, a handful of departments had missing data for 2020. When available, the most recent employee count since 2010 was used (231 departments). If a department did not have an employee count after 2010, that department's employee count was marked as missing (29 departments). Income and race/ethnicity data was provided by the 2019 American Census Survey. 148 counties were missing for median income of Black households and 55 counties were missing for median income for Hispanic households. Police departments selected for the study are associated with governments smaller than counties. However, it is unclear exactly which population each department would interact with. Accordingly, county-level data was deemed an appropriate choice for describing the economic and racial/ethnic character of a department's local area.

Relevant department characteristics are shown in Table 1. The Column 2 shows the mean value for the different characteristics for departments that received emails from White-male putative identities. Columns 3-7 give the difference between the White-male mean value and the other putative identities. Table 1 confirms that treatment was successfully randomized across the most obvious department characteristics relevant to this study. Only one of the differences across the columns is statistically significant at the 10% level (Pop. % Black (county-level)).

[12] During the police department email address collection process, 117 departments with populations less than 7,500 were incidentally included. These departments were included in the study.

Table 1: Balance Table

| **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 | 66,650 | -27.51 *(1,326)* | 71.09 *(1,328)* | -534 *(1,318)* | 338.5 *(1,321)* | -98.16 *(1,324)* |
| Median Income Black HH | 47,700 | -1,902 (1,469) | 500 (1,483) | -173 (1,457) | 635 (1,466) | 763 (1,473) |
| Median Income Hispanic HH | 53,180 | -676 (1,040) | -332 (1,042) | -683 (1,032) | -270 (1,033) | 475 (1,039) |
| Median Income White HH | 73,260 | 955 (1,448) | 681 (1,450) | -12 (1,439) | 338 (1,442) | 354 (1,446) |
| % Pop. in poverty | 0.12 | 0.0039 (0.0036) | 0.0040 (0.0036) | 0.0012 (0.0035) | -0.0011 (0.0035) | 0.0021 (0.0036) |
| % Black pop. in poverty | 0.22 | 0.0090 (0.0076) | -0.0007 (0.0077) | -0.0020 (0.0076) | -0.0061 (0.0076) | 0.0033 (0.0076) |
| % Hispanic pop. in poverty | 0.19 | -0.0004 (0.0056) | 0.0025 (0.0057) | 0.0013 (0.0056) | 0.000048 (0.0056) | -0.0016 (0.0056) |
| % White pop. in poverty | 0.09 | 0.0006 (0.0027) | 0.0008 (0.0027) | -0.0006 (0.0027) | -0.0017 (0.0027) | 0.0002 (0.0027) |
| **Population** | | | | | | |
| Local government pop. | 49,160 | 4,100 (12,760) | -1,080 (12,780) | 13,030 (12,680) | -10,370 (12,710) | 1,660 (12,740) |
| Sqrt. local government pop. | 170 | 9.74 (10.68) | 3.15 (10.70) | 8.71 (10.62) | -8.07 (10.64) | 5.59 (10.67) |
| Pop. % Black (county-level) | 0.10 | 0.0157* (0.0084) | 0.0019 (0.0085) | 0.0064 (0.0084) | 0.0007 (0.0084) | 0.0084 (0.0084) |
| Pop. % Hispanic (county-level) | 0.14 | -0.0021 (0.0111) | 0.0072 (0.0111) | 0.0037 (0.0110) | 0.0016 (0.0111) | 0.0033 (0.0111) |
| Pop. % White (county-level) | 0.69 | -0.0199 (0.0148) | -0.0145 (0.0148) | -0.0132 (0.0147) | -0.0023 (0.0147) | -0.0167 (0.0148) |
| Pop. % rural (county-level) | 0.21 | -0.0219 (0.0151) | -0.0129 (0.0151) | -0.0069 (0.0150) | 0.0061 (0.0150) | -0.0043 (0.0151) |
| **Department size (# of employees)** | | | | | | |
| Total employees | 128 | -0.87 (42.84) | -12.61 (42.94) | 47.77 (42.66) | -21.54 (42.78) | -7.96 (42.72) |
| Total officers | 104 | -1.35 (36.16) | -11.61 (36.24) | 39.99 (36.01) | -22.02 (36.11) | -8.88 (36.06) |
| Total civilian employees | 24 | 0.48 (7.62) | -1.00 (7.63) | 7.78 (7.58) | 0.48 (7.61) | 0.92 (7.60) |

*Standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 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 231 departments. 148 observations were missing for median income of Black households and 55 observations were missing for median income for Hispanic households. Emails were undelivered because the police department email address was unrecognized or police departments blocked the emails.

# 3 Results

## 3.1 Summary Stats

The first emails went out on Monday, June 27th and the last emails were sent out on Wednesday, August 31. In total 2,134 police departments were emailed. Table 2 breaks down the outcome of the emails. Of the 2,134 emails sent, 37 (denied or failed emails) were not included in the analysis.

Table 2: Emails categorized by outcome.

| **Email Category** | **Total** | **Percent of Total** |
|---|---|---|
| Sent | 2,134 | 100 |
| Response | 1,413 | 66.21 |
| No Response | 682 | 31.96 |
| Multiple Response | 207 | 9.70 |
| Denied | 13 | 0.61 |
| Failed | 24 | 1.12 |

As mentioned, emails were sent out in batches over the course of 10 weeks to reduce the chances of current events influencing police department response behavior. In Appendix **??**, Figure **??** depicts the response rate for all identities by week and Figure **??** breaks the weekly response rates down by putative identity. The figures suggest that, at least in the 10 weeks of the experiment, response behavior did not change considerably.

## 3.2 Main results

### 3.2.1 Overview

The main outcome for this study is a binary indicator for whether or not a police department responded to an email. For a department to be recorded as having responded, that department must reply within 4 weeks (28 days). Emails from departments that are automatically generated confirmations that the treatment email was received is not recorded as a response. Further discussion of alternative definitions of a response outcome variable are discussed below. The primary focus for this study concerns the differences in police department response behaviors to White putative identities and non-White (Black and Hispanic) putative identities. Accordingly, the omitted identity in the analysis is White or White male. [13] The present study uses fixed effects for the week an email was sent and the state of the department to account for treatment being stratified on week and state.

#### 3.2.2 General Results

Table 3 reports the most general analysis of differences in response rates across putative identities. Column 1 of Table 3 compares department response rates for emails with Hispanic putative identities (Hispanic emails) and Black putative identities (Black emails) to the mean response rate for emails with White putative identities (White emails). The response rate for White emails is 74.82%. Compared to the White email response rate, the response rate for Black emails is -10.38 [-16.39, -04.36] percentage points lower and the response rate for Hispanic emails is -10.62 [-15.70, -5.54] percentage points lower—both estimates are statistically significant at the 1% level.[14] Column 2 compares department response rates for emails with female putative identities to the mean response rate for emails with male putative identities (66.02%). The estimates from column 2 show that females, on average, were 2.23 [-3.93, 8.40] percentage points more likely to receive a response. However, the difference is not statistically significant. Column 3 includes all variables from column 1 and column 2. From column 3 the response rate for Black emails (-10.39 [-16.39, -4.39] percentage points) and Hispanic emails (-10.66 [-15.79, -5.53]) are slightly lower than estimates from column 1 and are still significant at the 1% level. Similarly, the response rate for female emails ( 2.35 [-3.68, 8.38]) are slightly higher than the estimate from column 2 and is not statistically significant at the 10% level. Explicitly testing for differences in explanatory power between model 1 and model 3 reveal that they are not statistically significantly different. However, the inclusion of female in model 3 shifting the Black and Hispanic email response rates further negative suggests that more can be learned by estimating separate coefficients for all six putative identities.

[13] Approximately 30 emails were blocked as spam by department email servers or failed to send due to incorrect department email addresses. Fortunately, the blocked and failed emails had an equal distribution across the putative identities.

[14] 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.

Table 3: General Results

| Dependent Variable: | | Response | |
|---|---|---|---|
| Model: | (1) | (2) | (3) |
| *Variables* | | | |
| Black | -0.1038*** | | -0.1039*** |
| | (0.0307) | | (0.0306) |
| Hispanic | -0.1062*** | | -0.1066*** |
| | (0.0259) | | (0.0262) |
| Female | | 0.0223 | 0.0235 |
| | | (0.0315) | (0.0308) |
| Reference group mean | | | |
| White | 0.7482 | | 0.7482 |
| Male | | 0.6602 | 0.6602 |
| *Fixed-effects* | | | |
| Week | Yes | Yes | Yes |
| State | Yes | Yes | Yes |
| *Fit statistics* | | | |
| Observations | 2,094 | 2,094 | 2,094 |
| $R^2$ | 0.05987 | 0.04939 | 0.06050 |
| Within $R^2$ | 0.01162 | 0.00059 | 0.01228 |

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

### 3.2.3 Interacted and Weighted Results

Table 4 breaks down response rates across the six different putative identities (3 race/ethnicity categories $\times$ 2 gender categories). The omitted identity is *White male*. This choice is made (1) so that estimates compare the groups that are commonly discriminated against (non-White and female) to the group commonly given preferential treatment (White male) and (2) for the easiest interpretation of results as the *White male* identity has the highest response rate (75.78) among the six putative identities. Column 1 gives the percentage point differential in response rates for the remaining five putative identities compared to the White male putative identity response rate. Column 2 reveals that response rates for Black males (-13.94 [-21.33, -6.56]) and Hispanic males (-15.00 [-21.94, -8.06]) are statistically significant at the 1% level and are are the lowest among all the putative identities. In comparison, the response rates is -9.70 [-17.48, -1.92]) for Black females and is -9.29 [-17.57, -1.00] for Hispanic females. The estimates, respectively, are statistically significant at the 5% and 10% level. 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 not statistically significant (p value = 0.3119), while the response rates for Hispanic males and Hispanic females are statistically significant (p value = 0.0035). The response rate for White females is -2.94 [-12.27, 6.40] percentage points lower, but is not statistically significant at the 10% level. White females are the only female putative identity to have a lower response rate than the male complement within each race/ethnicity grouping.

Table 4: Main Results

| Dependent Variable: | Response | |
|---|---|---|
| Model: | (1) | (2) |
| *Variables* | | |
| White × Female | -0.0294 | -0.0625 |
| | (0.0476) | (0.0542) |
| Hispanic × Male | -0.1500*** | -0.1490*** |
| | (0.0354) | (0.0408) |
| Hispanic × Female | -0.0929* | -0.1419** |
| | (0.0423) | (0.0512) |
| Black × Male | -0.1394*** | -0.1715*** |
| | (0.0377) | (0.0451) |
| Black × Female | -0.0970** | -0.1017*** |
| | (0.0397) | (0.0251) |
| *Weights* | | |
| Sqrt of local pop. | No | Yes |
| *Fixed-effects* | | |
| Week | Yes | Yes |
| State | Yes | Yes |
| *Fit statistics* | | |
| Observations | 2,094 | 2,094 |
| $R^2$ | 0.06211 | 0.07222 |
| Within $R^2$ | 0.01397 | 0.01616 |

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

Column 2 of Table 4 weights the observations by the square root of the local population.[15] Weighting

[15] The square root of 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 common method of logging the populations would reduce the disparity between populations too much. The log of Los Angeles's population is approximately 15, which comparatively similar to the log of the median local population ($log(18000) \approx 9.8$).

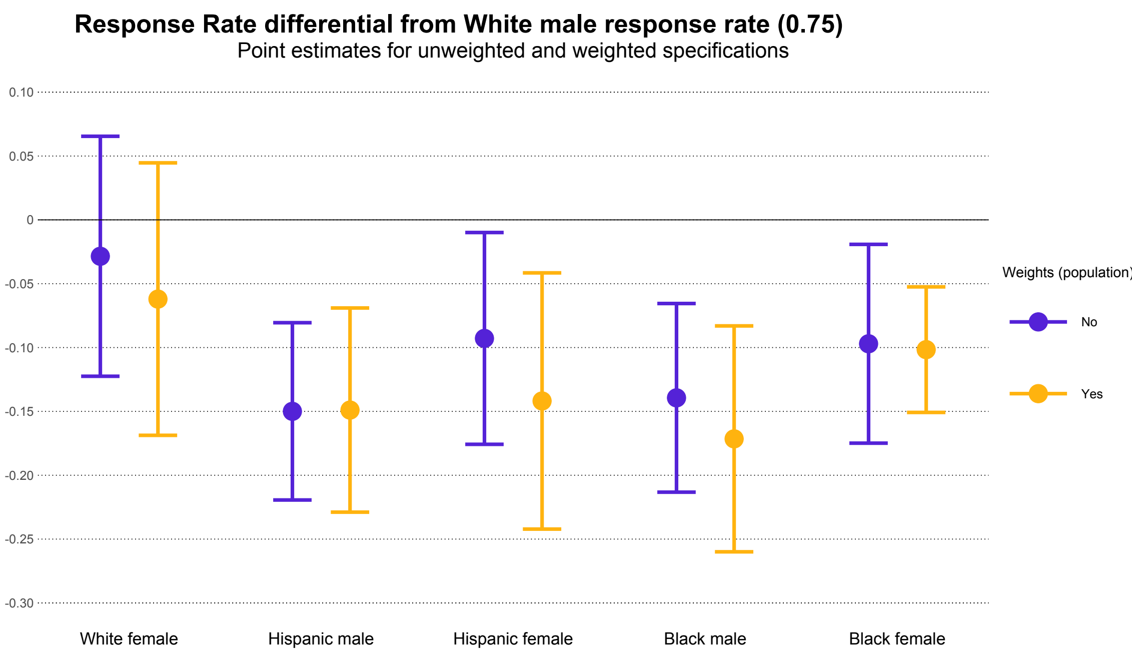



Figure 1: Response rate differentials from the response rate for White male identity from the other putative identities. The mean response rate for White male identities is 75%. Purple indicates unweighted estimates and gold indicates weighted estimates. Weights are determined by local population size for a department.

by local populations increases the disparity in the response rate for White males and all the other putative identities, expect for Hispanic males. The response rate for White females in column 2 (-6.25 [-16.88, 4.37]) is more than double the estimate from column 1, but remains statistically insignificant at the 10% level. Response rates for Black females (-10.17 [-15.08, -5.25]) and Hispanic females (-14.19 [-24.21, -4.16]) both increase in magnitude and statistical significance. The response rate for Black males (-17.15 [-26.00, -8.31]) becomes the putative identity with the lowest response rate. The response rate for Hispanic males (-14.90 [-22.89, -6.90]) decreases marginally in magnitude, but remains statistically significant at the 1% level. In contrast to column 1 of Table 4, 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 (p value = 0.0861), while the response rates for Hispanic males and Hispanic females *are not* statistically significant (p value = 0.8545).

#### 3.2.4 Sample split by Department Size

A potentially important factor in response behavior for a department is the number of employees that the department has at its disposal. Table 5 displays the results of models with the same specifications as Table 4 when the data is split into "small departments" and "big departments". The size of the department is determined by the median number of total employees for the departments included in the present study. Response rates are lower for White females, Black males and Black females in big departments compared to small departments. The White female response rates for small departments are 1.62 [-11.84; 15.08] (unweighted) or 2.10 [-12.72; 16.93] (weighted) percentage points *higher* than White male response rates. However, the White female response rates for big departments are -7.03 [-21.71; 7.65] (unweighted) or -9.81 [-24.93; 5.30] (weighted) percentage points *lower* than White male response rates. The unweighted differential response rate for Black males is more than twice as small for big departments (-18.92 [-36.25; -1.59]) as small departments (-8.92 [-16.64; -1.20]). When the observations are weighted by population, the differential response rate for Black males from big departments (-21.87 [-38.09; -5.65]) is almost three times as small as the response rate from small departments (-7.89 [-14.49; -1.28]). In contrast, the response rate for Hispanic males increases with department size, increasing from -17.55 [-26.63; -8.48] to -13.08 [-25.31; -0.84] when unweighted by population size, and increasing from -17.86 [-28.21; -7.51] to -14.43

[-26.25; -2.60] when weighting by population size. The most dynamic putative identity response rate is for Hispanic females. The unweighted results show that response rate *increases* with departments size from -8.80. [-19.40; 1.80] to -7.30 [-21.00; 6.39] percentage points, but when weighting by population size the response rate decreases with *decreases* with department size from -8.51 [-20.03; 3.01] to -14.79 [-29.95; 0.36] percentage points. Column 3 and Column 6 show the results of explicitly testing for differences in putative identity response rate estimates across sample sizes. Three of the differences are significant at the 10% level: Hispanic male response rates for the weighted and unweighted results, and the Black female response rate for the weighted results. In the other cases, the change in point estimates were matched with larger standard errors.

Table 5: Sample Split by Department Size

| Dependent Variable: | | | Response | | | |
|---|---|---|---|---|---|---|
| | Unweighted | | | Weighted by Population | | |
| Model: | Small Dept. (1) | Big Dept. (2) | **P value** | Small Dept. (3) | Big Dept. (4) | **P value** |
| *Variables* | | | | | | |
| White × Male mean | 0.747 | 0.765 | | | | |
| White × Female | 0.0162<br>(0.0595) | -0.0703<br>(0.0649) | 0.5605 | 0.0210<br>(0.0655) | -0.0981<br>(0.0668) | 0.5523 |
| Hispanic × Male | -0.1755$^{***}$<br>(0.0401) | -0.1308$^{**}$<br>(0.0541) | > 0.0000 | -0.1786$^{***}$<br>(0.0458) | -0.1443$^{**}$<br>(0.0523) | >0.0000 |
| Hispanic × Female | -0.0892$^{*}$<br>(0.0469) | -0.0730<br>(0.0605) | 0.1183 | -0.0851<br>(0.0509) | -0.1479$^{*}$<br>(0.0670) | 0.3045 |
| Black × Male | -0.0892$^{**}$<br>(0.0341) | -0.1892$^{**}$<br>(0.0766) | 0.6592 | -0.0789$^{**}$<br>(0.0292) | -0.2187$^{**}$<br>(0.0717) | 0.8172 |
| Black × Female | -0.0882<br>(0.0496) | -0.1039<br>(0.0569) | 0.1611 | -0.0813<br>(0.0516) | -0.1132$^{***}$<br>(0.0337) | 0.06056 |
| *Weights* | | | | | | |
| Sqrt of local pop. | No | No | | Yes | Yes | |
| *Fixed-effects* | | | | | | |
| Week | Yes | Yes | | Yes | Yes | |
| State | Yes | Yes | | Yes | Yes | |
| *Fit statistics* | | | | | | |
| Observations | 1,052 | 1,014 | | 1,052 | 1,014 | |
| $R^2$ | 0.08835 | 0.09749 | | 0.09078 | 0.11283 | |

Continued on next page

Table 5 – continued from previous page

| | | | | |
|---|---|---|---|---|
| Within $R^2$ | 0.01864 | 0.01648 | 0.02000 | 0.02028 |

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

# 4 Discussion

## 4.1 Results

### 4.1.1 General Results and Interacted Results Discussed

The results of this field experiment present strong evidence of racially biased policing. When aggregated across genders, compared to White emails, the response rate for Hispanic and Black emails are both 10 percentage points lower and statistically significant at the 1% level (Table 3). The question of gender biased policing is slightly more nuanced. Comparing response rates for all males to response rates for all females, it appears that the police are slightly more likely to respond to requests from females, thought the difference is not statistically significant. However, when response rates for each race/ethnicity-gender putative identity are compared, a different story emerges (Table 4). Perhaps unsurprisingly, White male identities receive the highest rate of responses–a reversal of the female identities receiving preferential treatment result. It is the low response rates for Hispanic males and Black males that drive the lower response rates for all males. Even though, in aggregate, there is not a statistically significant difference in male and female response rates, Hispanic females and Black females response rates are 9 percentage points lower than White males and significant at the 10% and 5% level, respectively. Comparing the results of Table 3 to Table 4 reveals that the *intersection* of race/ethnicity and gender is an important 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 putative 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, a crime is not being committed. Consequently, whether it is due to racist stereotypes, historic crime patterns, or a different mechanism, Black and Hispanic males are not treated the same way as their White counterparts.

An alternative explanation to the discrepancy across the six identities could be that departments hypothesize that the nature of the complaint might be different or that the likelihood of a formal complaint being pursued differs across groups. For instance, research suggest that police are more likely to use force with non-White citizens (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 coming from Hispanic and Black males are more likely to concern excessive use of force from one of their officers, and that type of complaint is more damaging to the department than other types of complaints.

In studies that examine racial discrimination, the question of statistical discrimination versus racial bias often comes up. In the case of this study, statistical discrimination on the grounds of different levels of criminal activity has already been addressed. However, another common refrain is that the discriminating against Black or Hispanic populations might be a symptom of these populations more frequently having lower socioeconomic statuses (SES) and police departments are discriminating on SES rather than race. For instance, Giulietti et al. (2019) take measures to separate the two in their correspondence study. However, if researchers are unable to disentangle the effect of minority status and lower SES on discrimination then in practice why does it matter? Impact over intent. Bertrand and Duflo (2017) argue that it is possible that police who disproportionately target non-White groups are engaging to some degree in both statistical targeting and biased policing (Bertrand and Duflo (2017)). Tilcsik (2021) argues that the idea of statistical discrimination "can lead people to view social stereotyping as useful and acceptable and thus help rationalize and justify discriminatory decisions."

#### 4.1.2 Weighted and Department Size Results Discussed

Police department response rates might also be driven by other factors. Size of a department (number of employees) and the population that a department serves could potentially affect response rates from departments. For instance, bigger departments might be able to 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. Departments that serve larger populations could mean that departments have more requests to fulfill. On the other hand, if departments serve smaller populations they might be more suspicious of the genuineness of the email they receive.[16]

The results reported in Table 3 and column 1 of Table 4 assign all departments equal weight. Consequently, the results shown in those tables are interpreted as response rates for an average police department. In other words, if a citizen were to contact a randomly selected police department, the response propensities listed in Tables 3 and 4 are the average response rates for the putative identities in this study. However, departments that serve larger populations interact with more citizens and are thus likely 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. Comparing Columns 1 and 2 in Table 4 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 3 is re-estimated, but this time interacting both *Hispanic* and *Black* with the five population bins. Figure 2 shows the results of this exercise. Figure 2 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 re-

[16] In the response from more than one department the police officer indicated that they checked the police logs and had no record of an interaction between one of their officers and a person that matched the name in the email they received.

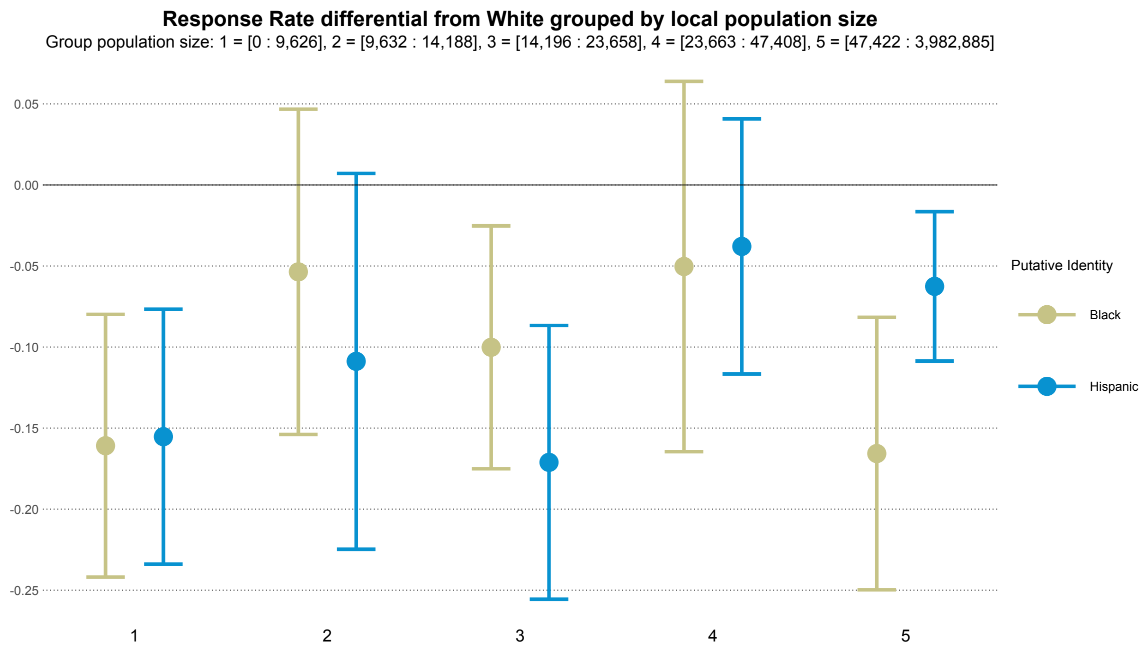



Figure 2: Response rates differentials for Black and Hispanic identities from White identities separated into bins determined by local population size.

sponse 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 4) 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.[17]

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 5 suggest that bigger departments are more likely to discriminate against Black identities and White female

[17]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.

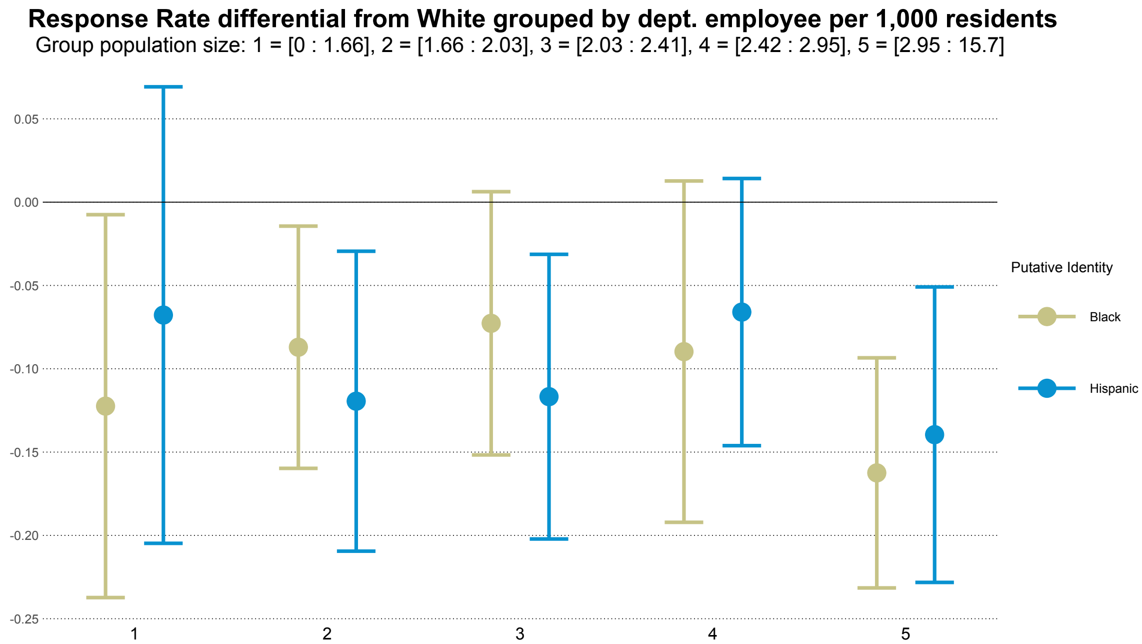



Figure 3: Response rates differentials for Black and Hispanic identities from White identities separated into bins determined by local population size.

identities, and discriminate less against Hispanic identities. When the results are weighted by population in Columns 3 and 4 in 5, the discrepancies grow, which makes sense given that agency size is correlated with population size.[18] To disentangle the effect of agency size and population size on response rate given their the two sizes correlations, the same exercise from Figure 2 is done. This time the quintiles are determined by the number of employees divided by local population size—a per capita measurement. Figure 3 displays the results of this exercise.

No clear pattern emerges for Hispanic or Black identity response rates. Compared to Figure 2, Figure 3 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

[18] One curious result of Table 5 is comparing the point estimate for Hispanic females from Column 2 and Column 4.

capable, in terms of employees per capita, to respond to requests are most likely to discriminate.

## 4.2 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 putative sender identities is worth exploring. For example, conditional on a department providing any response, do responses differ in their helpfulness and tone across putative 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 6: Response time and word count of response measured across putative 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) |

Continued on next page

Table 6 – continued from previous page

| | | |
|---|---|---|
| Black × Male | 4.518 | 8.778 |
| | (7.728) | (7.215) |
| Black × Female | -4.038 | 2.567 |
| | (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 6 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 putative identities and the other five putative identities. Table 6 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.

## 4.3 Accountability

The primary question of this study seeks to answer whether police departments discriminate on race, ethnicity or gender. The design of the study also emphasizes another important topic for policy makers interested in reforming police practices: accountability. By employing a correspondence study where citizens ask for assistance making a complaint against an officer, police departments are being asked to make a decision to respond to an inquiry based on a citizen's race/ethnicity and gender. However, they are also being assessed for their willingness to assist a citizen who is attempting to hold one of their officers accountable. In existing literature, the only prior correspondence study that includes law enforcement agencies is Giulietti et al. (2019). Their correspondence study interacts with a wide range of public institutions, where the list includes sheriffs' offices. In their study, the authors email the various public institutions with benign requests for relevant information, varying the putative identity of the citizen asking for information (where they use two distinctively black male names and two distinctively white male names). The authors find that these public institutions (ranging from public libraries to county clerks, in addition to sheriffs' 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 above 53% for White male emails and about 46% for Black male emails. Overall, these response rates are noticeably lower than average response rates in the present study. However, the difference in response rates by race are significantly smaller. One explanation for the difference is that Giulietti et al. (2019) target sheriff's offices instead of local police departments, and sheriffs' offices may face different expectations for accountability. However, an alternative explanation is that a simple request for general assistance is not treated 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 seem to be more responsive, but they also more evidence of racial/ethnic discrimination.

It should be noted that in both studies, Giulietti et al. (2019) and the present study, average response rates are quite 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 putative identity, White males, have a response rate of only 75%. By design, the complaints mentioned in the present study are fictitious. However, in reality, an officer behaving in such a way to warrant a citizen making an effort to file a formal complaint suggests potentially serious misconduct on the part of the officer. If only six out of ten citizens are able even to get assistance making a complaint, much less to have any action taken as a result of their complaint, it is difficult to consider citizen-initiated complaints about police officers as a reliable strategy for holding police officers accountable. This concern is amplified when groups of people who more often interact with police (i.e., non-Whites) are also less likely to be assisted in making a complaint.

### 4.4 Caveats

This study seeks to understand whether police departments tend to discriminate on the basis of race, ethnicity or gender. The results suggest that the average police department does. There are a few caveats to this study to keep in mind. First, police departments were selected randomly (see Appendix **??**). However, for a department to be eligible to be included in the study, it needed to have 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 certain type of department. Intuitively, it is likely 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%). It is difficult to make clear references from such a small sample, but 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. While the results clearly demonstrate that police departments in the United States have a higher propensity on average to respond to White emails as opposed to Black or Hispanic emails, the data provide only one observation per department. Thus is 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.[19] 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. In other contexts, for example, a police officer's decision to pull over a vehicle, the results of this study may or may not apply.

[19] Multiple requests to one department may have raised department suspicions about the inquiries.

## 4.5 Conclusion

This study uses a correspondence study to establish strong evidence for the existence of biased policing in the United States. From a sample of over thousand police departments, departments were 10 percentage points more likely to respond to emails from White putative identities than to emails from Black or Hispanic putative identities. When the racial/ethnic putative identities were interacted with gender, White male putative identities had the highest response rates and Black male and Hispanic male putative identities had the lowest response rates—respectively, 13.94 and 15 percentage points lower than the White male response rates. Further analysis of the responses that were received and a follow-up RCT that seeks to examine within-department behavior would would greatly complement the striking findings of this study. The low overall response rates and large heterogeneity in response rates across putative identities are concerning. Low response rates suggest that police departments are resistant to accountability. Furthermore, response rates being significantly lower for emails from non-White identities, suggest that departments are especially unwilling to engage with communities of color. Although the literature remains inconclusive about the existence of biased policing, the results of this study suggest that biased policing does exist, and it is imperative to identify biased policing in other contexts.

## 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.

Bandes, S. A., Pryor, M., Kerrison, E. M., and Goff, P. A. (2019). The mismeasure of Terry stops: Assessing the psychological and emotional harms of stop and frisk to individuals and communities. *Behavioral Sciences and the Law*, 37(2):176–194.

Bertrand, M. and Duflo, E. (2017). Field Experiments on Discrimination a aLaura Stilwell and Jan Zilinsky provided excellent research assistance. We thank Abhijit Banerjee for comments. We are particularly grateful to Betsy Levy Paluck, our discussant, for her detailed and thoughtful revie. pages 309–393.

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

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

Chalfin, A. and McCrary, J. (2017). Criminal deterrence: A review of the literature.

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

Cheng, C. and Long, W. (2018). Improving police services: Evidence from the French quarter task force. *Journal of Public Economics*, 164:1–18.

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.

Evans, W. N. and Owens, E. G. (2007). COPS and crime. *Journal of Public Economics*, 91(1-2):181–201.

Feigenberg, B. and Miller, C. (2021). Would Eliminating Racial Disparities in Motor Vehicle Searches have Efficiency Costs? *The Quarterly Journal of Economics*, 137(1):49–113.

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.

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.

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.

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.

Goel, S., Rao, J. M., and Shroff, R. (2016). Precinct or prejudice? Understanding racial disparities in New York city's stop-and-frisk policy. *Annals of Applied Statistics*, 10(1):365–394.

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

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

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.

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

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.

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.

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.

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.

Shoub, K., Christiani, L., Baumgartner, F. R., Epp, D. A., and Roach, K. (2021). Fines, Fees, Forfeitures, and Disparities: A Link Between Municipal Reliance on Fines and Racial Disparities in Policing. *Policy Studies Journal*, 49(3):835–859.

Singla, A., Kirschner, C., and Stone, S. B. (2020). Race, Representation, and Revenue: Reliance on Fines and Forfeitures in City Governments. *Urban Affairs Review*, 56(4):1132–1167.

Smith, M. R., Rojek, J. J., Petrocelli, M., and Withrow, B. (2017). Measuring disparities in police activities: a state of the art review.

Su, M. (2020). Taxation by Citation? Exploring Local Governments' Revenue Motive for Traffic Fines. *Public Administration Review*, 80(1):36–45.

Su, M. (2021). Discretion in Traffic Stops: The Influence of Budget Cuts on Traffic Citations. *Public Administration Review*, 81(3):446–458.

US Department of Justice Civil Rights Division (2015). Investigation of Ferguson Police Department. Technical report.

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

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.

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

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

# A Online Appendix: Putative 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)

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.

# B 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.[20] 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.

[20]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".

# C Online Appendix: Example Email

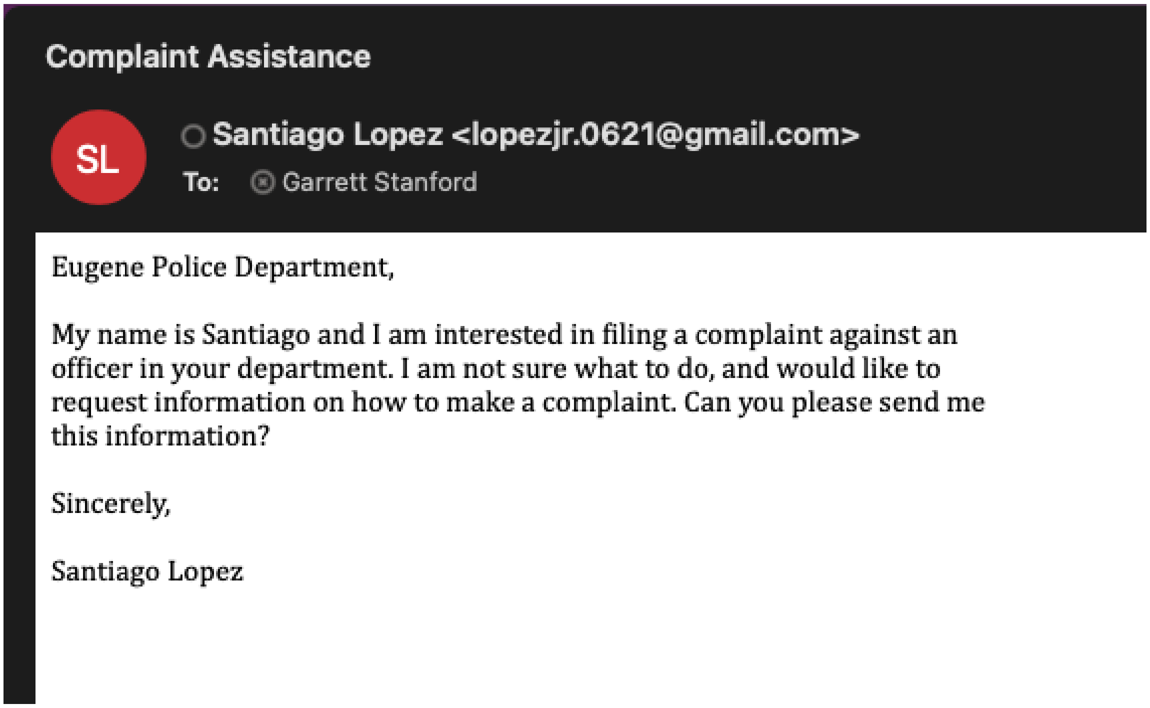
## C.1 Email Text

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. [21]

[21] 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

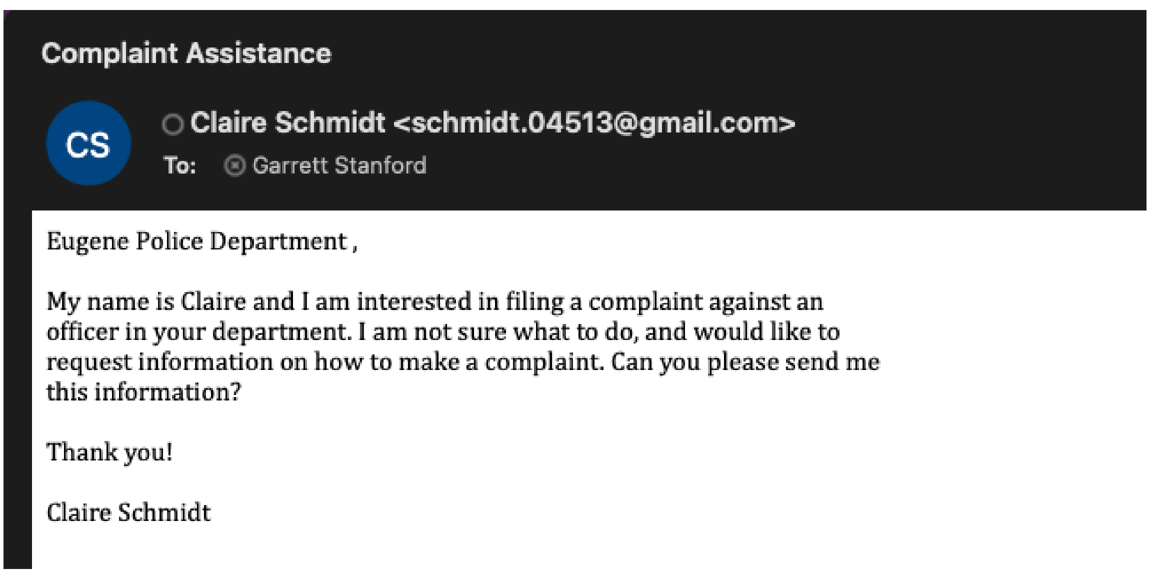
Eugene Police Department,

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

# D 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.

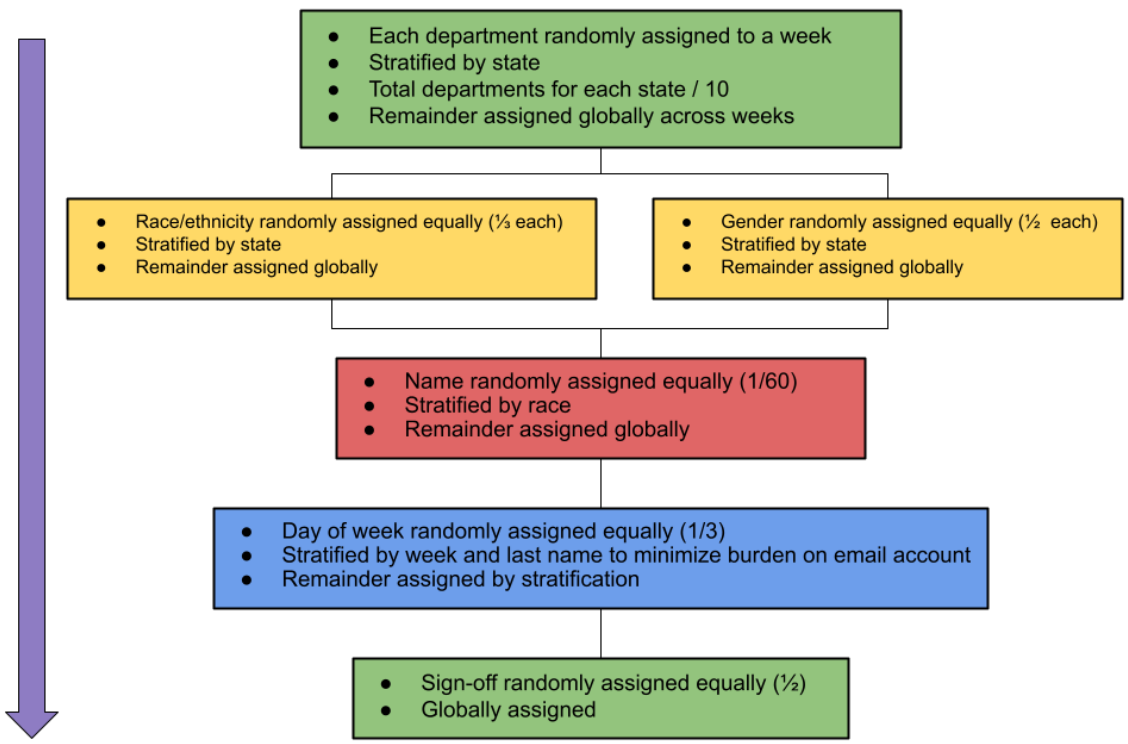


Table B1: Distribution of Race, Ethnicity and Gender identity assignment by state

| 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 |

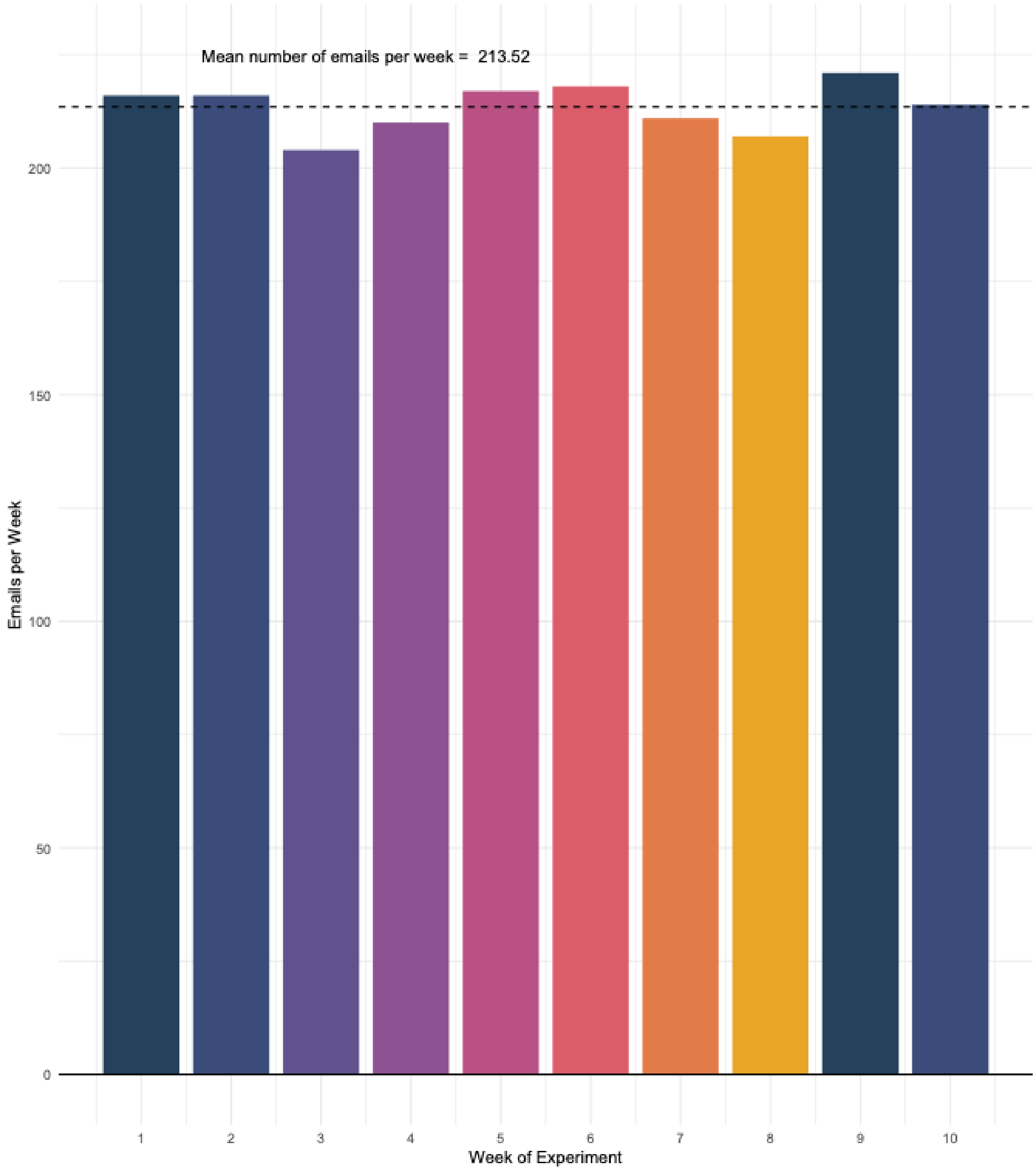



Figure B1: Emails sent by week.

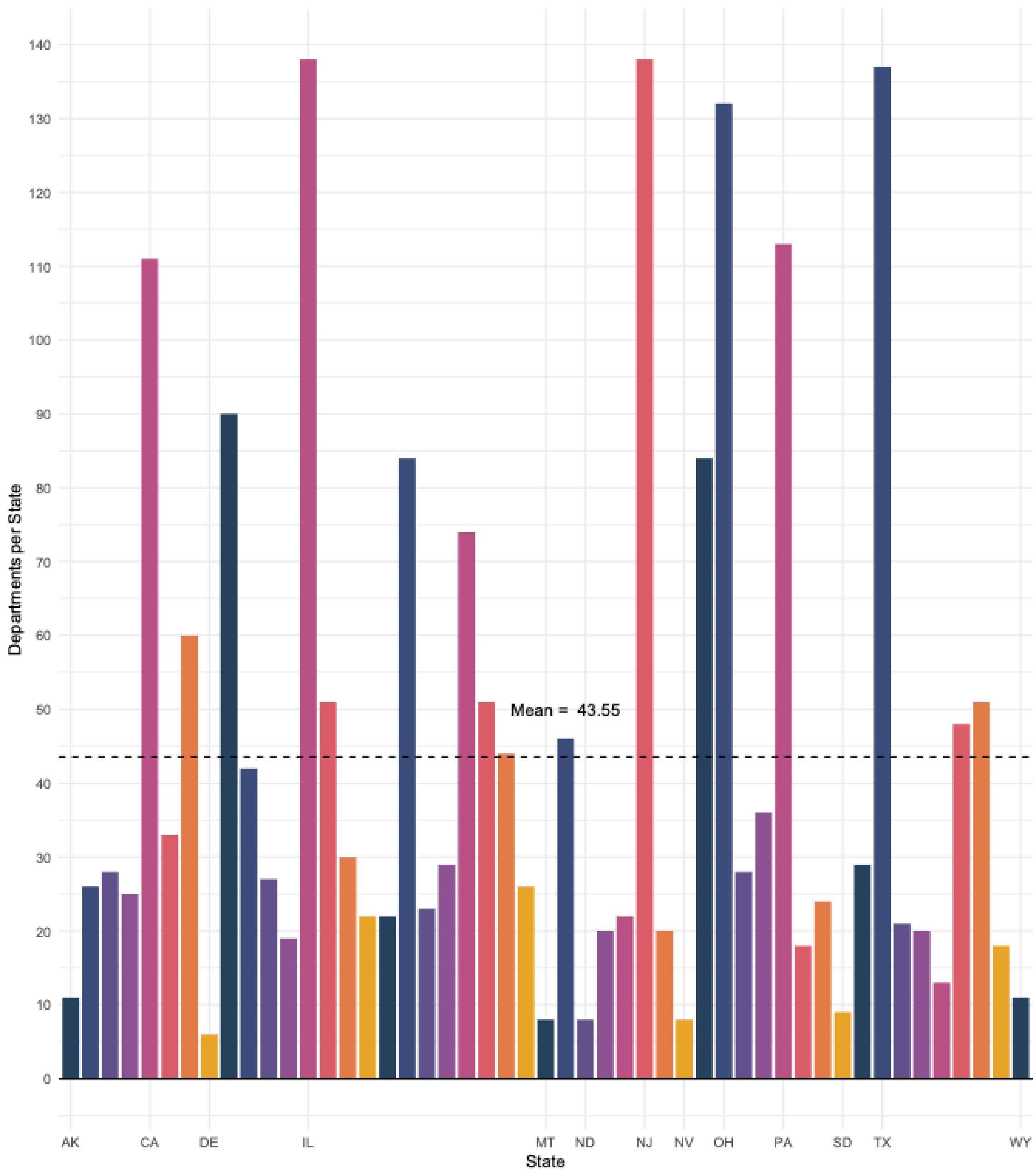



Figure B2: Departments by state included in the study.

# E Online Appendix: Police Department Selection

# 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.[22] 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

[22] 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.).