

The Black-White Recognition Gap in Award Nominations*

Nayoung Rim[†]

Roman Rivera[‡]

Andrea Kiss[§]

Bocar Ba[¶]

November 16, 2020

There is substantial evidence showing racial bias in firms' hiring decisions, but less is known about bias in career recognition. We construct a dataset of police award nominations to measure bias against minority employees. Exploiting quasi-random variation in supervisor assignment, we find that white supervisors are less likely to nominate black officers than white or Hispanic officers. Increased supervisor-officer interaction reduces the gap but not completely. We conduct an online experiment and find respondents are less likely to acquire information about black officers vis-a-vis non-black officers. Our findings suggest bias in career recognition is important for the black-white promotion gap.

JEL Codes: J71, M51, J45, J48

Keywords: racial bias, police officers, award nominations, supervisors

*A prior version of this working paper circulated as “In-group Bias and the Police: Evidence from Award Nominations”. We would like to thank Peter Arcidiacono, Pat Bayer, Sandra Black, Kerwin Charles, Rob Garlick, Rachel Greenspan, Kareem Haggag, Damon Jones, Trevon Logan, Corinne Low, Bentley MacLeod, Conrad Miller, Robert Moffitt, Samuel Myers, Jr., and the Huettel Laboratory for helpful comments. We thank Sam Stecklow, the Invisible Institute, and Craig Futterman for help with the data. Weizhao Sun provided excellent research assistance. We thank the Duke Economics Department, the U.S. Naval Academy, and the Quattrone Center at Penn Law School for generous financial support. IRB approval for this experiment was received from the U.S. Naval Academy (#USNA.2019.0040-PADS) and Duke University (#2020-0338). The experiment was pre-registered in the AEA RCT Registry, AEARCTR-0005929. The views expressed herein do not reflect the position of the Chicago Police Department or the U.S. Naval Academy. All errors are our own.

[†]Corresponding author. U.S. Naval Academy, Department of Economics. Email: rim@usna.edu

[‡]Columbia University, Department of Economics. Email: r.g.rivera@columbia.edu

[§]Duke University, Department of Economics. Email: a.kiss@duke.edu

[¶]University of California-Irvine, Department of Economics. Email: bba@uci.edu

1 Introduction

For decades, a goal of public policy has been to reduce racial disparities in the labor market.¹ The economics literature has largely focused on firms’ hiring decisions because of the ability to experimentally examine hiring (Bertrand and Mullainathan, 2004; Kessler et al., 2019; Neumark et al., 2019). Less is known about racial bias in career recognition and progression, which may arguably be more important for the lack of diversity in upper-management positions and, ultimately, the racial wage gap.

An important question for eliminating discrimination and racial gaps in career outcomes is whether supervisors choose to interact with and acquire information about minority colleagues. We examine this question in the context of the second largest police department in the US, where supervisors do not necessarily observe the officer’s day-to-day activities but are required to evaluate the officer’s performance annually. Because supervisors do not directly monitor officers, they must exert effort to gather information on officers when it comes to the annual evaluation. This may manifest in a racial gap in career recognition and progression. Although our application focuses on law enforcement, this organizational structure (autonomous workers operating within a hierarchical organization) is common across all industries.

We construct a novel panel dataset of all Chicago Police Department (CPD) officers between 2009 and 2015 containing detailed personnel information on use of force, arrests, and misconduct—crucial information in an empirical study of bias in the workplace. Using supervisor nominations for departmental awards, we examine whether white supervisors are less likely to acquire information about and nominate their minority officers. We focus on award nominations rather than wage and promotion because nominations are subjective evaluations of officers’ performance. In contrast, wages in the CPD vary only by experience

¹In regards to understanding the source of discrimination in the labor market, the economics literature has coalesced around two main explanations: taste-based discrimination and belief-based or statistical discrimination. See Lang and Kahn-Lang Spitzer (2020) for a survey of the literature.

due to a union contract, and promotions are largely determined by a written test.² Further, awards are an important measure of career recognition and are used in important decisions related to career advancement, such as performance evaluations, merit promotions, and overtime pay.

Our identification strategy exploits two institutional features of the CPD that allows us to obtain plausibly causal estimates of the black-white recognition gap. First, officers are assigned a new supervisor every January. We confirm as-good-as-random assignment by analyzing supervisor-officer assignments and find that officers do not sort to supervisors based on work performance measures.³ Second, all officers must be evaluated annually by their supervisor, and the quarter of evaluation is randomized across officers.⁴ Under the assumption that supervisors are more likely to interact with and gather information about officers closer to the evaluation period, this institutional feature allows us to exploit the randomly assigned evaluation quarter and estimate the causal impact of an interaction, which would normally be endogenous.

Building off of Rim et al. (2020), we first document that an officer’s arrest record has a positive and significant impact on the likelihood of being nominated for an award. But, conditional on the number of arrests, black officers are less likely to be nominated for an award relative to white officers, and the negative black-white gap widens with the number of arrests. Specifically, the black-white gap nomination gap among officers with five or more arrests is -2.6 percentage points—an effect size of 82.6 percent, as the mean nomination probability for a white officer by a white supervisor is 3.1 percent. The black-white recognition gap is largest among less experienced supervisors and among supervisors with a higher share of

²Seventy percent of promotions to sergeant are determined by a written exam, while 30 percent may be based on nominations by higher-ranking officers. These nominations take into consideration an officer’s qualifications, such as the number of awards received. Further, promotions are rare because the sergeant exam is not offered on a regular basis, thereby limiting the opportunities for police officers to be promoted.

³In particular, we may be concerned that more-productive white officers and/or less-productive black officers sort to white supervisors. In this case, we would see a negative black-white nominations gap from white supervisors even in the absence of racial bias.

⁴The evaluation must be held in the quarter prior to the quarter that the officer joined the CPD, and the quarter in which officers join CPD is determined by lottery number.

black employees, and does not vary much by the number of supervisees. We also find that supervisors are more likely to nominate all officers in the quarter of evaluation, suggesting that statistical discrimination may be at play (Altonji and Pierret, 2001).⁵ However, this increase in nomination likelihood is larger for white officers relative to black officers and persists only for white officers, suggesting that statistical discrimination is not the only explanation for this racial disparity. If lack of information were the only reason for the racial gap in award nominations, we would expect that full information in the evaluation quarter would eliminate the racial gap.

Because our observational CPD data do not capture detailed interactions between supervisor and officer, we conduct an online experiment to measure the review process in the nomination decision. We ask Amazon Mechanical Turk (MTurk) workers to evaluate officer profiles and nominate one for an award. In addition to being able to experimentally examine the nomination decision process, the online experiment allows us to generalize our findings to a broader evaluator group than Chicago police supervisors.

In one task, MTurk workers choose between a black officer and a non-black officer. Although officer performance levels are randomly chosen, MTurk workers are 5.6 to 8.8 percentage points less likely to nominate black officers over white officers. In another task, officer profiles display only demographic information and workers must mouse over the profile to reveal full information about the officer. We monitor mouse movements across the screen and find that black officers are less likely to be moused over. This result is more pronounced when workers are choosing among three white officers and one black officer. We do not see similar patterns for Hispanic officers when workers are choosing among three white officers and one Hispanic officer. Taken together, our findings suggest that the negative black-white nomination gap is due to a lack of interest about black officers.

Our paper relates to the literature on social networks in the workplace. Prior research documents the importance of gender homophily on career outcomes (Cullen and Perez-

⁵Altonji and Pierret (2001) argue that as firms learn more about their employees, statistical discrimination should decrease.

Truglia, 2020; Sarsons, 2019; Zeltzer, 2020). This paper expands the literature by examining the importance of race homophily in career recognition. For example, we find that CPD supervisors and MTurk workers, both groups that are mostly white, are less likely to gather information on minority officers, leading to a racial disparity in award nominations. Furthermore, CPD supervisors continue nominating their white officers but not their black or Hispanic officers after the annual evaluation. These findings are consistent with Bartoš et al. (2016), which finds that employers are less likely to open and read resumes from minority candidates. Our findings also speak to the literature that documents benefits of same-race matching on racially disparate outcomes. There is mounting evidence that race-matching leads to better outcomes, such as in education (Carrell et al., 2010; Gershenson et al., 2018; Kofoed and McGovney, 2019) and health (Alsan et al., 2019), but less is known about why. Our results suggest that networks formed through race homophily are important for success in the workplace.

Our paper is also similar to other papers that find that discrimination may arise because biased managers interact less with minorities (Glover et al., 2017). Our experimental evidence finds that black officers are less likely to be moused over, particularly when they are in a pool with three white officers. Additionally, MTurk workers spend more time evaluating black officers’ profiles but are not more likely to nominate them. These findings are consistent with studies that find that minorities are less likely to be acknowledged for their work (Hengel, 2019; Sarsons, 2020) and a strand of literature that establishes the existence of bias among managers and work colleagues (Bertrand and Mullainathan, 2004; Bohren et al., 2019; Egan et al., 2018; Giuliano et al., 2009; Glover et al., 2017; Sarsons, 2019).⁶ By analyzing the black-white recognition gap among police officers, our paper links this literature to studies on racial disparities in law enforcement.

With respect to law enforcement, our study adds to the growing research that is uncov-

⁶Hengel (2019) and Sarsons (2020) find that female minorities are less likely to be acknowledged for their work.

ering racial bias in policing.⁷ Prior studies largely use data on officer-initiated encounters, which may be biased because they do not include the universe of all possible police interactions (Knox et al., 2020). Notable exceptions exploit randomly assigned officer dispatches to 911 calls or to investigate automobile crashes (Weisburst, 2018; West, 2018). Similarly, our paper bypasses the truncated data problem by focusing on supervisor nominations of quasi-randomly assigned officers. Additionally, we ask whether racial bias on the part of officers carries over to their colleagues, a question that was previously unanswered due to a lack of detailed personnel data.

We begin the rest of the paper with a short description of CPD’s organizational structure and the awards nomination process (Section 2). Section 3 describes our data collection efforts and presents summary statistics on our CPD analysis sample. Section 4 tests the identifying assumption that officers do not sort to supervisors based on work performance measures. We present results using observational CPD data in Section 5. Section 6 discusses the experimental evidence. We discuss the evidence for statistical discrimination and taste-based discrimination in Section 7. We conclude with a discussion of the policy implications for law enforcement agencies in Section 8.

2 Background

2.1 Basic Facts about CPD’s Structure

After passing a written exam, all Chicago Police Department candidates are placed on an eligibility list according to a randomly assigned lottery number and called off in lottery order to enroll in police academy. Upon graduation from Police Academy, Police Officers begin their career in one of the 25 geographic districts spanning the city of Chicago.⁸ These

⁷See, for example, Ajilore and Shirey (2017); Antonovics and Knight (2009); Anwar and Fang (2006); Bacher-Hicks and de la Campa (2020); Close and Mason (2006); Cunningham and Gillezeau (2018); Goncalves and Mello (2020); Hoekstra and Sloan (2020); Horrace and Rohlin (2016); Knowles et al. (2001); Mason (2007); Nix et al. (2017).

⁸Between 2012-2014, three districts were dissolved leaving 22 geographic districts.

initial assignments are generally outside the officer’s control, with the exception of a small number of officers who received academic and other distinctions in the Academy (Police Accountability Task Force, 2016).⁹

Police Officers are supervised by Sergeants in their district. Daily responsibilities for sergeants include participating in roll call, supervising criminal investigations (protecting the scene, establishing the perimeter, etc.), and ensuring officers carry out their responsibilities.¹⁰ Every year, sergeants conduct performance evaluations of their assigned supervisees.¹¹ To assist supervisors with performance evaluations, an electronic database called the Performance Recognition System (PRS) tracks exceptional or adverse behavior related to job performance. Information is entered by Human Resources staff, and supervisors have the ability to monitor and track information in PRS, though it is uncertain whether any actually do in practice (U.S. Department of Justice, 2017, p. 111-112).

CPD patrol officers work a rotational schedule, where they rotate their off-days each week. Therefore, officers are not necessarily assigned to work on the same days as their supervisors who conduct the annual performance evaluation (U.S. Department of Justice, 2017, p. 108).

2.2 CPD Awards Nomination Process

The Chicago Police Department distributes department awards to recognize the accomplishments, performance, and service of its Department members. In addition to highlighting officers’ accomplishments, awards are used in important decisions related to career advance-

⁹When vacancies occur, officers may bid for district transfers. Successful bidders are chosen based on their qualifications and seniority.

¹⁰Section III.A., Employee Resource E05-05, retrieved from <http://directives.chicagopolice.org> and Appendix A, CPD Sergeant Written Assessment Study Briefing 2013, retrieved from https://www.chicago.gov/content/dam/city/depts/dhr/general/CPD_Sergeant_Assessment_Study_Briefing_2013.pdf.

¹¹The average supervisor conducts 7.8 evaluations a year. The median number of evaluations (supervisees) is seven.

ment, such as performance evaluations¹², merit promotions¹³, and overtime pay¹⁴.

We focus on award nominations in this paper because they are subjective evaluations of officers' performance.¹⁵ The importance of subjective evaluations is underscored in the Police Accountability Task Force (2016), which states that—despite the test-based promotional process in the CPD—98 percent of CPD officers felt promotions were due to connections instead of merit. This sentiment is consistent with a growing literature that documents the importance of mentoring and networks in the workplace (Beaman et al., 2018; Cullen and Perez-Truglia, 2020).

More practically, we focus on award nominations because of institutional facts that restrict our study of wage and promotion in the CPD. Officer base salaries are set by a pay schedule determined solely by experience. Therefore, there is no racial disparity in pay by construction. Additionally, promotions are rare in the CPD because they depend on the number of vacancies, which occur when a higher-ranking officer retires or dies, and because the sergeant exam is not offered on a regular basis; almost ten years passed between the two most recent sergeant promotion exams (Police Accountability Task Force, 2016, p. 140). As it is difficult to study racial disparities in officer career progression with these traditional measures, one contribution of this paper is its access to data on award nominations.

There are 33 departmental awards, which range in their competitiveness. Most awards require a nomination process. Nominations may originate from any higher-ranking officer, including one's supervisor.¹⁶ Our analysis focuses on nominations by officially assigned supervisors to leverage the institutional feature that requires assigned supervisors to conduct

¹²Chicago Police Department, Career Development Directive, Employee Resource E05-01, Section IV.H., retrieved from <http://directives.chicagopolice.org>.

¹³Merit promotions are a more subjective selection process that rely on a variety of officer qualifications including the number of awards, vis-a-vis test-based promotions. See, Section III.E.2, Employee Resource E05-05, retrieved from <http://directives.chicagopolice.org>.

¹⁴The number of awards is a statistically significant predictor of overtime pay. An additional award last year is correlated with an additional \$206.78 in overtime pay this year, which is more than four times the effect of an additional arrest last year (\$46.99). These estimates control for years of experience.

¹⁵We choose not to examine award receipt because those are determined by an external Awards Committee, which may add an additional layer of bias in the decision process.

¹⁶Nearly 90 percent of nominations for police officers are from sergeants. Thirteen percent of all award nominations are from an officer's assigned supervisor.

performance evaluations at random quarters in the calendar year.

Officers may be nominated for a single award per incident, and nominations must be submitted within 45 days of the incident.¹⁷ There is no restriction on the number of times an officer may be nominated, as long as the nominations are for different incidents. Supervisors are also not restricted in the number of award nominations they are allowed to submit.

3 Data

This section describes administrative police records and district-level crime information that are used for our empirical analysis. We first describe the data sources and the linked analysis dataset. Then, we provide descriptive statistics of Police Officers in the Chicago Police Department between 2009 and 2015.

3.1 Police Officer Data

Administrative records and information on sworn Chicago Police Department members were obtained by Freedom of Information Act requests through a collaboration with Invisible Institute. In order to connect different datasets, officers are first identified within a dataset using the available unique characteristics, such as name, appointed date, birth year, and race, and then matched with identified officers in different datasets.

Demographics Data on officer race, sex, birth year, and appointment date are obtained from aggregated data, using the most common observation across datasets.¹⁸ Officer rank is taken from salary data provided by the Chicago Department of Human Resources (DHR), covering 2002 to 2017.

¹⁷There are a few exceptions to this. The Carter Harrison/Lambert Tree Medal, 100 Club of Chicago Valor Award, Superintendent's Award of Valor, Police Blue Star Award, and Police Blue Shield Award may be awarded to officers who received other departmental awards for the same incident (Chicago Police Department, Department Organization Directive, Special Order S01-01, retrieved from <http://directives.chicagopolice.org>).

¹⁸Not all demographic information is complete in each file, so an aggregation of demographic variables across multiple files is necessary. Over 99% of officers are matched to a unique gender, race, and appointment date.

Supervisors This dataset provides information about the supervisor who conducted each officer’s annual evaluation between 2009 and 2017. Our analysis focuses on those at the rank of Police Officer, meaning their supervisors are at the rank of Sergeant. In this paper, the term “supervisor” refers to a Sergeant who is officially assigned to conduct a Police Officer’s annual evaluation in a given calendar year.

Awards The awards dataset provides information on all department award nominations between 2004 and 2017. The dataset includes the award name, the individual being nominated, the requester, request date, and the final status of the nomination (approved, deleted, or denied).¹⁹ We consider all performance awards that are open to all sworn Department members and require a supervisor’s nomination.²⁰ After these restrictions, our analysis considers 18 awards. Appendix Table A1 provides a description of these awards.

Unit Assignment Historical unit assignment data lists all units to which an officer was assigned since the beginning of his or her career, as well as start- and end-dates in each unit. We focus our analysis on Police Officers assigned to the 25 geographic districts.

Arrests The arrests dataset contains information on all arrests made by Department members. The dataset includes detailed information about the subject, crime, and arrest location and time. These data cover 2001-2017 but arrest day and month are only provided from 2010 onwards. For arrests made in 2009, we use the date the subject was released from the local police station as a proxy for the arrest date.²¹ For our analysis, we use total officer arrests as well as arrests in three aggregated crime categories: violent crime, property crime, and “non-index” crime. The Federal Bureau of Investigation classifies violent and property crimes as “index crimes” because they are more serious offenses.²² Non-Index crimes capture

¹⁹An award may be deleted for various reasons, including: the form was not filled out correctly; supporting evidence was not included; or the nomination does not meet the eligibility requirements of the award. This differs from an award denial, which means the officer did not win the award. Very few awards (2.4 percent) are deleted.

²⁰Most awards are open to all Department members. One example of an exception is the Thomas Wortham IV Military and Community Service Award, which is awarded to current or former members of the U.S. Armed Services.

²¹In 96.9% of cases, the release date is on the same day or the day after the arrest date, and 100% of release dates are within four days of the arrest.

²²Violent crimes are crimes related to violence, such as murder and assault. Property crimes are crimes

crimes that are not related to violence or property, such as municipal code violations, traffic violations, warrants, drugs, prostitution, gambling, etc.²³

Complaints The complaints data contain all recorded allegations of misconduct filed against an officer from 2000 to 2016. Allegations may originate from the public or from other officers in the department.

Tactical Response Reports Data on officer use of force come from 2004-2016 Tactical Response Reports (TRR). Officers are required to file a TRR if they used any force while performing their duties. A TRR filing requirement can be triggered by three things: the subject’s actions; the officer’s actions; or a subject who is injured or alleges injury resulting from the officer’s use of force option. CPD publishes a Use of Force Model, which provides guidelines on the appropriate level of force to be used in response to a subject’s actions and levels of resistance. Using the Use of Force Model as a guide, we classify officer force options into two broad categories of “weak use of force” and “strong use of force.” Weak use of force includes force mitigation efforts, such as verbal direction and tactical positioning (which involve no physical touch), and control tactics, such as escort holds and wristlocks. Strong uses of force involve elevated levels of force that are generally intended to enact harm on or injure the subject.²⁴ The data only report use of force against adult persons. Appendix Table A2 outlines force options and our classification.

Sample restrictions. To construct a complete dataset on all officers in the Chicago Police Department, we require that officers receive a salary from DHR and appear in the unit assignment dataset. We focus on years 2009 to 2015 to maximize overlap across the different datasets. We further restrict our sample to officers at the rank of Police Officer who are always assigned to a geographic district²⁵ and officer-supervisor relationships that lasted

related to property, such as burglary and motor vehicle theft.

²³A comprehensive list of crime categories can be found at http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html.

²⁴Strong use of force may or may not use weapons. Examples of strong use of force without weapons are take-downs, kicks, and punches. Examples of non-lethal weapons are chemical weapons and long-range acoustic devices. Examples of lethal weapons are tasers, batons, and firearms.

²⁵We remove the three districts that closed between 2012-2014 (13, 21, and 23) from our analysis sample because we do not have crime statistics for these districts.

for 12 months. Our final analysis dataset has 6,518 Police Officers and 1,284 supervisors. In terms of the outcome variable, we consider nominations for 18 awards that require a supervisor’s nomination and is open to all Department members.

3.2 Crime Data

We use crime data from the Chicago Data Portal (<https://data.cityofchicago.org>), which contains reported incidents of crime that occurred in the City of Chicago since 2001. The dataset contains the primary type of crime, the date, location, and whether the crime led to an arrest. We construct monthly crime rates²⁶ for each district, separately for total crimes, property crimes, and violent crimes.

3.3 Summary Statistics

This section provides descriptive statistics of Police Officers in our analysis sample. From Table 1, we see that most officers are male (73.7 percent) and white (46.4 percent), but blacks and Hispanics are also well-represented (23 to 27 percent). In fact, these three racial groups make up nearly 97 percent of our sample. The average CPD officer in our sample joined the force in 2000 at age 30. This indicates that at the start of our analysis dataset (year 2009), the average officer had been on the force for 9 years.

Relative to Police Officers, the racial makeup of supervisors in our analysis sample is more homogeneous. About 81 percent of supervisors are male, and 70 percent are white. Blacks and Hispanics each make up around 14 percent of supervisors. At the start of our analysis dataset, the average supervisor had worked for 17 years or 8 years longer than the average Police Officer.

Table 2 presents racial differences in various work measures. The first row is the probability of being nominated for an award in a particular month. For example, the average

²⁶Crime rate is defined as the total number of reported incidents of crime divided by the population and multiplied by 1000.

Table 1: Summary Statistics

| Sample: | Police Officers | Supervisors |
|-----------------|-----------------|-------------|
| Male | 73.7% | 80.8% |
| Race | | |
| White | 46.4% | 69.7% |
| Black | 26.8% | 14.7% |
| Hispanic | 23.2% | 14.0% |
| Asian | 3.1% | 1.6% |
| Native American | 0.4% | 0.1% |
| Birthyear | 1970.3 | 1965.3 |
| Start Year | 2000.0 | 1992.2 |
| Observations | 6,518 | 1,284 |

Source: CPD analysis sample.

officer has a 2.5 percent chance of being nominated in a given month, which equates to about a 30 percent chance of being nominated in a given year. Whites and Hispanics have slightly higher than average likelihoods at 3 percent and 3.2 percent, respectively, while the likelihood for black officers is half the sample average (1.3 percent). The black-white difference is statistically significant at the 1 percent level.

The second row in Table 2 lists the number of monthly complaints. The average officer receives about 0.04 complaints in a given month, equating to about 1 complaint every two years. This statistic is similar across race.

The remaining rows in Table 2 depict the number of monthly arrests by arrest type. For example, the average officer makes 1.8 arrests every month. White and Hispanic officers are slightly over this average at 2 and 2.2 arrests, respectively, while black officers are below this average at 1.2 arrests. The black-white difference equates to 10 fewer arrests a year. This is statistically significant at the 1 percent level.

When comparing summary statistics for the different types of arrests, we see that the black-white difference in total arrests is driven by arrests for non-index crimes, which make up around 65 percent of all arrests. Here, the difference is about -0.70 arrests per month or 8.4 fewer arrests per year and is statistically significant at the 1 percent level.

Table 2: Racial and Gender Differences in Work Measures

| Sample: | All Officers | White Officers | Black Officers | Hispanic Officers | B-W Difference (p-value) | H-W Difference (p-value) |
|---------------|-----------------|-------------------|-------------------|----------------------|-----------------------------|-----------------------------|
| Nominated | 2.5% | 3.0% | 1.3% | 3.2% | -1.7 (0.000) | 0.2 (0.016) |
| Complaints | 0.04 | 0.04 | 0.04 | 0.04 | 0.00 (0.937) | 0.00 (0.075) |
| Total Arrests | 1.82 | 2.04 | 1.19 | 2.16 | -0.85 (0.000) | 0.12 (0.000) |
| Violent | 0.37 | 0.37 | 0.31 | 0.42 | -0.06 (0.000) | 0.05 (0.000) |
| Property | 0.27 | 0.29 | 0.20 | 0.30 | -0.09 (0.000) | 0.01 (0.017) |
| Non-Index | 1.19 | 1.38 | 0.68 | 1.44 | -0.69 (0.000) | 0.07 (0.000) |
| Drug | 0.31 | 0.37 | 0.14 | 0.41 | -0.23 (0.000) | 0.03 (0.000) |
| Traffic | 0.12 | 0.15 | 0.06 | 0.16 | -0.09 (0.000) | 0.01 (0.002) |
| TRR filings | 0.05 | 0.05 | 0.03 | 0.06 | -0.02 (0.000) | 0.00 (0.039) |
| Observations | 250,872 | 111,876 | 70,572 | 59,148 | | |

Source: CPD analysis sample.

Notes: This table lists monthly summary statistics for 6,518 police officers. Sample is at the officer-month level. Non-index arrests include arrests for non-property and non-violent crimes. B-W Difference reports the percentage-point difference between black officers and white officers. H-W Difference reports the percentage-point difference between Hispanic officers and white officers. p-values are the p-value from a t-test of a difference in means.

Although the data reveal a disparity in number of arrests, we caution the reader from jumping to the conclusion that black officers are less productive than white or Hispanic officers. Arrests are not a comprehensive measure of policing quality and may even be a biased measure (Owens et al., 2018). For example, a comparison of white and black officers' arrest records would suggest that black officers are less productive than white officers. However, a study by Harvey and Mattia (2020) finds that police departments that increased their share of black officers subsequently reduced black crime victimization. Similarly, female officers have fewer arrests than male officers but Miller and Segal (2018) finds that increasing the number of female police officers decreased the number of intimate partner homicides and increased the number of reports of domestic violence in the U.S. These outcome measures, which are important measures of social welfare, are not captured by arrests nor would they appear on an officer's record.

Another example is to consider drug and traffic arrests, which are presumably proactive in that they are more likely to have originated from an officer-initiated incident. This

classification of “proactive arrests”, which allows for greater officer discretion, can also be seen as a delineation between appropriate and inappropriate uses of police authority.²⁷ In Table 2, we see that white officers are about 2.4 to 2.6 times more likely than black officers to arrest someone for drugs or traffic violations. In contrast, the black-white difference for more serious crimes, like violent crimes, is economically small at -0.06 arrests a month. Similarly, Ba et al. (2020) examine daily patrol assignments of CPD officers and find that black officers make fewer stops and arrests and use force less often than their white colleagues. This disparity is driven by a decreased focus on discretionary contact, such as stops for “suspicious behavior”. These facts suggest that although it is important to control for work measures in our analysis, we should not automatically interpret differences in overall arrests as differences in policing quality.

4 Key Identifying Assumption

This section outlines the empirical strategy to examine whether a black-white recognition gap exists in award nominations. We use the assignment to a new supervisor at the start of a calendar year to approximate random assignment of an officer’s race to a supervisor.²⁸ If we believe the as-good-as-random assignment, then we can simply compare nomination rates of a white officer vs. a black officer assigned to a given supervisor to test for the existence of racial bias. This identification strategy is similar to Chetty et al. (2014), which develops a quasi-experimental method by exploiting student exposure to teachers of varying value-added levels that is induced through teacher turnover. In this paper, we use officer exposure to supervisors of different races that is induced through annual re-assignment of

²⁷We borrow this term and classification from Worden et al. (2013). We do not know whether an arrest stemmed from an incident that the officer initiated on his or her own authority, but we assume that drug and traffic arrests are more likely to have stemmed from officer-initiated traffic stops as compared to arrests for violent crimes. Importantly, proactive arrests should be considered as a very noisy measure of quality policing. For example, Worden et al. (2013) analyzed the impact of a police agency’s early intervention system, which aims at monitoring and managing police misconduct among officers who exhibit patterns of problematic behavior, and found that it lowered the number of proactive arrests with little impact on productivity.

²⁸About 96% of officers are assigned to a supervisor in January of each calendar year.

supervisors to officers.

Several facts suggest that officers do not game the supervisor assignment system. First, about 89 percent of all officer-supervisor relationships are new relationships. Second, 78 percent of all supervisor relationships between 2009 and 2015 lasted exactly one year. Although the vast majority of supervisor relationships last one year, because it is not a totality, we may be concerned that some officer-supervisor relationships may have been arranged outside of the random assignment system. Therefore, we restrict our analysis sample to all supervisor-officer relationships that last one year in order to minimize the number of endogenously formed supervisor relationships. In the next section, we test whether officers are as-good-as-randomly assigned to supervisors in the data.

4.1 Exogeneity of Supervisor Assignment and Officer Performance

Throughout the paper, we want to interpret any change in nomination likelihood when white supervisors are assigned white officers relative to when they are assigned black officers as a causal effect of officer race. The key assumption is that minority officers were not systematically assigned to white supervisors in years when officer performance would have been particularly low for other reasons. For example, if high-performing white officers and low-performing black officers sort to white supervisors, then we would see a negative black-white nomination gap. This may appear to be bias against black officers by white supervisors, but in reality it would be the result of sorting of police officers based on work performance measures. We argue that this sorting concern is mitigated in our setting due to the as-good-as-random assignment of supervisors every January.

One way to examine the validity of this argument is to test whether officers of different races are differentially likely to be assigned to a supervisor of a given race. Because supervisor assignments occur every January, if there is supervisor-sorting, then we would expect last year's performance measures to be correlated with this year's supervisor assignment. Again, the sorting concern is that white supervisors are assigned to higher-performing white officers

relative to black officers. Because officers are assigned to supervisors at the Department level, we use all patrol officers assigned to a supervisor rather than the analysis sample that is restricted to officers whose supervisor assignment lasted one year.

We estimate the following regression model:

$$Y_{it} = \beta_0 + X'\beta + Z'\alpha_1 + (B_i \times Z')\alpha_2 + (H_i \times Z')\alpha_3 + (A_i \times Z')\alpha_4 + (N_i \times Z')\alpha_5 + e_{it} \quad (1)$$

where Y_{it} is equal to 1 if officer is assigned to a white supervisor in year t and 0 otherwise. X is a vector of baseline controls, such as officer sex, birth year, tenure at the end of the calendar year, district assignment at the start of the year, year fixed effects, and district-year fixed effects. Z is a vector of lagged work performance measures, such as the number of arrests (e.g., violent crimes, property crimes, and non-index crimes), complaints, TRR filings, and strong use of force ratio. Although we include a rich set of control variables, there may be unobservable characteristics that may affect the supervisor assignment process. For example, perhaps more aggressive police officers tend to sort to more aggressive sergeants. To mitigate this concern, we include a measure for the level of force. B_i is a binary indicator variable if the officer is black, H_i if Hispanic, A_i if Asian, and N_i if Native American. White officers are the reference group.

Table 3 presents estimates for α_2 , the black-white difference in lagged officer work measures, and for α_3 , the Hispanic-white difference in lagged officer work measures, when the dependent variable is a white supervisor (columns 1 and 2). As a robustness check, we also estimate equation (1) for black supervisors in columns 3 and 4.

The evidence suggests that white, black, and Hispanic police officers do not differentially sort to white supervisors based on their prior year's performance. Two of the three work measures that are statistically significant have trivial estimates that are essentially zero. The Hispanic-white difference in TRR filings is 0.018, and though this is significant at the 5 percent level, it represents a 3 percent increase, which is not economically significant (the

average officer has 0.56 annual TRR filings). We also conduct a joint F-test to see if the racial differences in lagged work performance measures are jointly significant in being assigned to a white supervisor. We find that the black-white difference—the focus of this paper—is not jointly significant (p-value = 0.896). The Hispanic-white difference is borderline significant with a p-value of 0.107.

In columns 3 and 4, we examine assignment to black supervisors. Violent-crime arrests has racial gaps that are statistically significant but trivial magnitudes. The black-white difference is -0.0069, and the Hispanic-white difference is -0.002. Considering that the average officer makes 4.1 annual violent-crime arrests, these estimates are not economically significant. In terms of the joint F-test, we find that the black-white difference and the Hispanic-white difference are both jointly significant.

Table 3: Sorting of Officers to Supervisors

| Dependent Variable: Coefficients for: | Supervisor is White | | Supervisor is Black | |
|--|-------------------------|-------------------------|--------------------------|-------------------------|
| | Black-White Diff. | Hispanic-White Diff. | Black-White Diff. | Hispanic-White Diff. |
| | (1) | (2) | (3) | (4) |
| Violent-crime Arrests | 0.00182 (0.00244) | 0.00320 (0.00198) | -0.00687*** (0.00184) | -0.00242** (0.00106) |
| Property-crime Arrests | -0.000293 (0.00269) | 0.000767 (0.00154) | 7.52e-05 (0.00214) | 2.60e-05 (0.000795) |
| Non Index-crime Arrests | -3.79e-05 (0.000390) | -0.000130 (0.000293) | 0.000468 (0.000346) | -0.000174 (0.000179) |
| Complaints | 0.00214 (0.0109) | 0.000370 (0.00884) | 0.0131 (0.00874) | 0.000524 (0.00438) |
| TRR Filings | 0.00747 (0.0107) | 0.0177** (0.00866) | -0.000203 (0.00885) | -0.00476 (0.00447) |
| Strong Force Ratio | -0.0229 (0.0260) | 0.0102 (0.0231) | 0.0207 (0.0209) | 0.00965 (0.0118) |
| p-value for joint F-test | 0.896 | 0.107 | 0.005 | 0.092 |
| Observations | 10,276 | | 10,276 | |

Source: CPD data.

Notes: This table reports the black-white difference and the Hispanic-white difference from a regression of officer race on having a white supervisor (columns 1 and 2) and having a black supervisor (columns 3 and 4). Additional controls include officer sex, birth year, tenure, district assignment, year fixed effects, and district-year fixed effects. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Another analysis is to compare average lagged annual work performance measures of officers assigned to white supervisors vs. black supervisors, separately by officer race. Table 4 reports average lagged performance measures by officer race and supervisor race. Panel A

lists mean lagged annual measures for white officers assigned to white supervisors in column 1 and to black supervisors in column 2. Panel B lists the same for black officers, and Panel C lists the same for Hispanic officers.

The table says that the average white officer assigned to a white supervisor this year had 4.6 violent-crime arrests last year and 3.96 property-crime arrests last year. The average white officer assigned to a black supervisor this year had 4.12 violent-crime arrests last year and 3.23 property-crime arrests last year. The third column lists the percentage-point black-white difference and p-value in parentheses from a t-test of a difference in means. Although some of the work measures have a statistically significant black-white difference, they are all economically small. The largest difference is in non-index crime arrests made by white officers assigned to white supervisors vs. black supervisors, but this is equal to 2.43 annual arrests when the annual average is over 20. Overall, these numbers suggest that officers assigned to white supervisors vs. black supervisors look very similar in terms of work performance.

Taken together, the results suggest that white supervisors are assigned to officers who are, on average, similar in terms of their work performance. They also provide an indirect test of random assignment of officers to white supervisors as we would not expect to see any correlation between lagged officer work measures and supervisor assignment if officers are randomly assigned to supervisors. Of course, we cannot test whether any unobservable traits are similar between white and black officers assigned to the same supervisor. For example, we may be concerned that, due to homophily, white officers may feel more comfortable than their black colleagues in opening up to their white supervisors. If this is the case, supervisors may be more informed of the achievements of white officers relative to black officers, leading to a black-white nomination gap. We address this potential concern by leveraging the randomized quarter of evaluation in [Section 5.2](#).

Table 4: Officer Work Measures and Supervisor Race, Comparison of Means

| | White Supervisor | Black Supervisor | B-W Difference (p-value) |
|----------------------------|---------------------|---------------------|-----------------------------|
| Panel A: White Officers | | | |
| Violent-Crime Arrests | 4.60 | 4.12 | -0.49 (0.003) |
| Property-Crime Arrests | 3.96 | 3.23 | -0.73 (0.0001) |
| Non-Index Crime Arrests | 20.58 | 23.01 | 2.43 (0.031) |
| TRR Filings | 0.61 | 0.59 | -0.02 (0.540) |
| Strong Force Ratio | 0.88 | 0.91 | 0.03 (0.257) |
| Complaints | 0.54 | 0.59 | 0.05 (0.161) |
| Panel B: Black Officers | | | |
| Violent-Crime Arrests | 3.80 | 3.25 | 0.54 (0.000) |
| Property-Crime Arrests | 2.67 | 2.08 | -0.59 (0.000) |
| Non-Index Crime Arrests | 10.11 | 9.14 | -0.97 (0.091) |
| TRR Filings | 0.38 | 0.30 | -0.07 (0.001) |
| Strong Force Ratio | 0.88 | 0.89 | 0.01 (0.624) |
| Complaints | 0.53 | 0.51 | -0.02 (0.302) |
| Panel C: Hispanic Officers | | | |
| Violent-Crime Arrests | 5.20 | 4.31 | -0.90 (0.0002) |
| Property-Crime Arrests | 4.06 | 3.02 | -1.04 (0.0001) |
| Non-Index Crime Arrests | 19.67 | 20.58 | 0.91 (0.537) |
| TRR Filings | 0.65 | 0.48 | -0.17 (0.002) |
| Strong Force Ratio | 0.91 | 0.97 | 0.05 (0.152) |
| Complaints | 0.58 | 0.54 | -0.03 (0.451) |

Source: CPD analysis sample.

Notes: This table reports mean lagged annual work measures for officers assigned to white supervisors and black supervisors. B-W Difference reports the percentage-point difference between black officers and white officers. p-values are the p-value from a t-test of a difference in means. Non-index arrests include arrests for non-property and non-violent crimes.

5 Results

5.1 Main Results

In this section, we examine whether an officer's arrest record affects the supervisor's likelihood of nomination and whether there are any differential effects for minority officers. That is, conditional on the officer's arrest record, are there racial differences in the probability

of nomination? The regression sample for this analysis is at the officer-month level. We estimate the following model, separately for white supervisors and black supervisors:

$$\begin{aligned}
Nom_{it} = & \beta_0 + \left(\sum_{c=1}^5 \mathbb{1}\{Arrests_{i,t-1} = c\} \times \beta_1^c \right) + \left(B_i \times \sum_{c=1}^5 \mathbb{1}\{Arrests_{i,t-1} = c\} \right) \beta_2^c \\
& + \left(H_i \times \sum_{c=1}^5 \mathbb{1}\{Arrests_{i,t-1} = c\} \right) \beta_3^c + \left(A_i \times \sum_{c=1}^5 \mathbb{1}\{Arrests_{i,t-1} = c\} \right) \beta_4^c \quad (2) \\
& + \left(N_i \times \sum_{c=1}^5 \mathbb{1}\{Arrests_{i,t-1} = c\} \right) \beta_5^c + X'\alpha + \tau_t + \varepsilon_{it}
\end{aligned}$$

where i denotes officer and t denotes month. Nom_{it} is equal to 1 if officer i was nominated for an award in month t and 0 if not. $Arrests_{i,t-1}$ is the number of arrests officer i made last month. We lag arrests because nominations must be submitted within 45 days of an incident. The reference category is zero arrests last month.

B_i is a binary indicator variable if the officer is black, H_i if Hispanic, A_i if Asian, and N_i if Native American. White officers are the reference group.

X is a vector of officer, supervisor, and district characteristics. Officer controls include officer's birth year, district assignment, tenure, and the number of complaints made against the officer. Supervisor controls include supervisor fixed effects and the share of black supervisees. District characteristics include overall crime rate and violent crime rate. All time-varying variables except for tenure, district assignment, and the share of black supervisees are lagged by one month. We also include fixed effects for year and month in τ_t . We estimate robust standard errors to account for heteroskedasticity introduced by the binary dependent variable.

The parameters of interest are β_2^c , which tells us how the black-white difference changes by the number of arrests, and β_3^c , which tells us how the Hispanic-white difference changes by the number of arrests. We expect β_1^c to be positive and increasing in the number of arrests. This is based on the fact that officers are nominated for an award based on an incident. If there is no racial disparity in award nominations, then β_2^c will be zero. A negative β_2^c

indicates that supervisors are more likely to nominate white officers (or, are less likely to nominate black officers), whereas a positive β_2^c indicates that supervisors are more likely to nominate black officers vis-a-vis white officers.

Table 5 reports estimates for β_2^c and β_3^c separately for white supervisors in Panel A and for black supervisors in Panel B. We report estimates for white officers in column 1, the black-white difference in column 2, and the Hispanic-white difference in column 3. There are increasing returns to having more arrests, with a marked increase for those with five or more arrests (column 1). Although we do not assert that arrests are an accurate measure of policing quality, we do the analysis this way because police departments seem to value and reward arrest quantity. It is interesting, therefore, that the return to having more arrests is less for black officers compared to white officers (Panel A, column 2). The black-white difference in nomination probability for officers with one arrest widens by 0.5 percentage-points compared to the black-white difference among officers with no arrests last month. This estimate is significant at the 5 percent level.

The black-white nomination gap widens even more as the number of arrests increases. Among officers with five or more arrests, the relative black-white difference widens by 2.4 percentage-points and is significant at the 1 percent level. It is informative to interpret this disparity in the context of racial differences in work performance. For example, black officers with 5 or more monthly arrests are at the 94th percentile of their distribution, while white officers are at the 81st percentile of their distribution. Yet, white supervisors are even *less* likely to nominate black officers over white officers compared to if both had zero arrests.

Table 5: Impact of Arrest Record on Nomination Likelihood by Officer Race

| Estimates for: | Outcome Variable: Nominated | | |
|---------------------------------|-----------------------------|--------------------------|---------------------------|
| | White Officer (1) | Black-White Gap (2) | Hispanic-White Gap (3) |
| Panel A: White Supervisors | | | |
| 1 arrest | 0.00913*** (0.00135) | -0.00496** (0.00193) | -0.000886 (0.00239) |
| 2 arrests | 0.0147*** (0.00190) | -0.00758*** (0.00278) | 0.000313 (0.00329) |
| 3 arrests | 0.0216*** (0.00260) | -0.0105*** (0.00408) | 0.000170 (0.00447) |
| 4 arrests | 0.0250*** (0.00326) | -0.00248 (0.00586) | -0.000702 (0.00566) |
| 5 or more arrests | 0.0566*** (0.00256) | -0.0235*** (0.00481) | -0.0156*** (0.00428) |
| Observations | | 171,094 | |
| Panel B: Black Supervisors | | | |
| 1 arrest | -0.000463 (0.00309) | 0.00354 (0.00354) | 0.00737 (0.00639) |
| 2 arrests | 0.00857* (0.00509) | -0.00640 (0.00570) | 0.000598 (0.00997) |
| 3 arrests | 0.0101 (0.00659) | -0.000213 (0.00800) | 0.00949 (0.0137) |
| 4 arrests | 0.00851 (0.00889) | -0.0151 (0.0102) | 0.00817 (0.0171) |
| 5 or more arrests | 0.0451*** (0.00762) | -0.00302 (0.0113) | -0.0222* (0.0125) |
| Observations | | 29,413 | |
| Mean Pr(Nom) for White Officers | | 0.031 | |

Source: CPD analysis sample.

Notes: This table reports estimates for the impact of an officer's arrest record on the probability of nomination by white supervisors (Panel A) and by black supervisors (Panel B). Each panel is a single OLS regression with estimates for white officers in column 1, the black-white difference in column 2, and the Hispanic-white difference in column 3. All estimates include supervisor, month, and year fixed effects, and control for officer birth year, tenure, district, lagged arrests, lagged complaints, lagged overall crime rate, lagged violent crime rate, and the share of black supervisees. The mean nomination probability of white officers by white supervisors is 3.1%. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We also examine whether white supervisors are less likely to nominate Hispanic officers, another racial minority in the Chicago Police Department (Panel A, column 3). The Hispanic-white difference is pretty trivial and not statistically significant until the five or more arrests category. Among officers with at least five arrests, the Hispanic-white gap in

nomination probability widens by 1.6 percentage-points and is significant at the 1 percent level. As Hispanics are more likely to be nominated than white officers in the zero arrest category, this means the Hispanic-white gap becomes negative.

When comparing between the two racial minority groups, the black-white difference is statistically significantly different from the Hispanic-white difference in the 1 to 3 arrests categories and not for those with 4 or more arrests. This suggests that white supervisors penalize black officers more than white or Hispanic officers among those with average arrest records (recall the average number of arrests is two), but favor white officers when comparing officers with more arrests.

Panel B reports estimates for officers assigned to black supervisors. None of the point estimates are significant, though this may be due to the fact that there are few black supervisors (190 compared to 893 white supervisors). However, the magnitudes of the point estimates are also very small. The one exception is that the Hispanic-white gap among officers with five or more arrests is weakly significant at -0.0222.

In Appendix Table A4, we estimate a version with officer fixed effects instead of supervisor fixed effects. The results are very similar. We also examine how the black-white nominations gap changes by supervisor characteristics (Appendix Figure A1). The gap is larger among supervisors with a higher share of black supervisees and among less experienced supervisors. It does not change much by the number of supervisees.

5.2 Information Acquisition

In this section, we examine a potential mechanism for the black-white nomination gap. The 2016 report by the Police Accountability Task Force found little stability in supervisor-officer relationships. Not only are supervisors reassigned every January, but officers may not be assigned to work with their officially assigned supervisor during the course of their shift. Second, personnel information does not necessarily get transferred to supervisors when officers switch assignments. Therefore, one potential explanation for why white supervisors

may be less likely to nominate black officers is because they are less likely to interact with them and, therefore, are less likely to be informed of their accomplishments (Glover et al., 2017). To test this theory, we exploit an institutional feature that randomizes the quarter in which officers are evaluated by their supervisor. Although there appears to be little interaction between officers and supervisors on a daily basis, we assume that the annual evaluation requires supervisors to acquire information about the officer’s work record.

All supervisors are required to conduct annual evaluations of their assigned officers, and this evaluation must take place during the quarter prior to the quarter in which the officer joined the Department. Appendix Table A3 lists the evaluation quarters and evaluation due dates by start month. For example, if an officer started his career in July (Q3), then his annual evaluation must take place in the second quarter of every calendar year. Because start dates are randomly determined by a lottery number, this means that the evaluation quarter is essentially randomly assigned across officers.²⁹

We exploit this institutional feature and compare nomination likelihoods of black vs. white officers assigned to white supervisors across quarters. Because the evaluation quarter is randomly assigned, this simple comparison allows us to isolate the effect of acquiring information. If a lack of information acquisition is the reason for a black-white nomination gap, then we would expect this to disappear in the quarter when supervisors are required to evaluate their assigned officers. For this analysis, the sample is at the officer-month level, and the regression model is:

$$\begin{aligned}
Nom_{it} = & \beta_0 + \sum_{q=-2}^3 \mathbb{1}\{EQ = q\} \delta^q + \left(B_i \times \sum_{q=-2}^3 \mathbb{1}\{EQ = q\} \right) \beta_1^q \\
& + \left(H_i \times \sum_{q=-2}^3 \mathbb{1}\{EQ = q\} \right) \beta_2^q + \left(A_i \times \sum_{q=-2}^3 \mathbb{1}\{EQ = q\} \right) \beta_3^q \\
& + \left(N_i \times \sum_{q=-2}^3 \mathbb{1}\{EQ = q\} \right) \beta_4^q + X' \alpha + \tau_t + e_{it}
\end{aligned} \tag{3}$$

²⁹After passing a written exam, all CPD candidates are placed on a eligibility list according to a randomly assigned lottery number and called off in lottery order to enroll in the police academy.

where i denotes officer and t denotes month. Nom_{it} is equal to 1 if officer i was nominated for an award in month t and 0 if not. B_i is a binary indicator variable if the officer is black, H_i if Hispanic, A_i if Asian, and N_i if Native American. White officers are the reference group.

The second term is a set of binary indicator variables for each quarter relative to the evaluation quarter, which is denoted as $EQ = 0$. The reference quarter is $EQ = -3$, or three quarters prior to the evaluation quarter. The coefficients δ^q tell us how nomination likelihoods for white officers change across quarters. If information acquisition is an important mechanism, then we expect it to be enhanced in the quarter that supervisors evaluate their officers (δ^0).

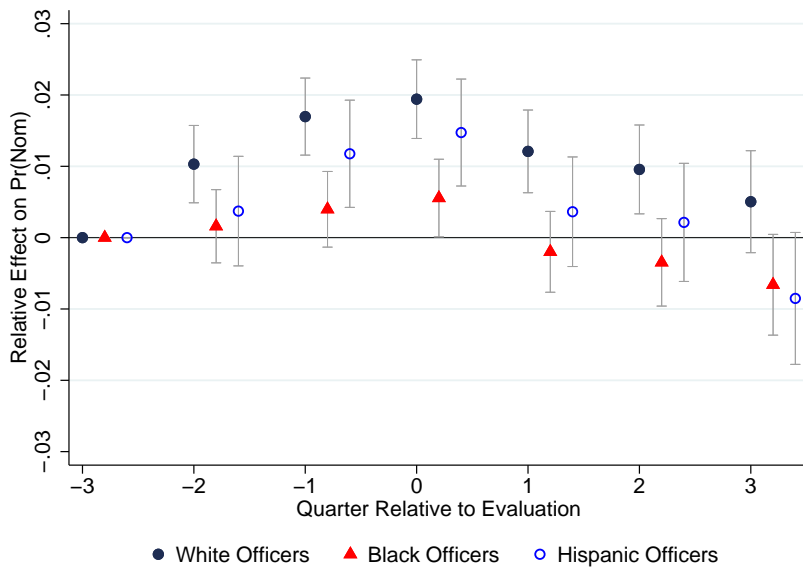
The third term in parentheses interacts the black indicator variable and the relative-quarter indicator variables. The coefficients β_1^q depict how the black-white nomination gap evolves relative to $EQ = -3$. If white supervisors are equally likely to nominate their black and white officers, then we expect β_1^q to be zero. Likewise, the coefficients β_2^q tell us how the Hispanic-white nomination gap evolves over time.

X is a vector of officer, supervisor, and district characteristics. Officer controls include officer’s birth year, district assignment, tenure, the number of arrests the officer made, and the number of complaints made against the officer. Supervisor controls include supervisor fixed effects and the share of black supervisees. District characteristics include overall crime rate and violent crime rate. All time-varying variables except for tenure, district assignment, and the share of black supervisees are lagged by one month. We also include fixed effects for year and month in τ_t . We estimate robust standard errors.

Figure 1 plots the estimates for δ^q separately for white officers, black officers, and Hispanic officers.³⁰ The hump-backed shape suggests that the nomination probability increases as we grow closer to the evaluation quarter then falls afterwards. This pattern exists for both white and black officers assigned to white supervisors, but the hump is less steep for black officers.

³⁰Specifically, we plot estimates for δ^q when the reference group is white officers. Then, we re-estimate equation (3) with black officers as the reference group and plot δ^q on the same graph.

Figure 1: Probability of Nomination by Quarter



Source: CPD analysis sample.

Notes: This graph shows how the probability of nomination changes by quarter relative to three quarters before the officer's evaluation, separately for white, black, and Hispanic officers. Sample is restricted to all officers assigned to a white supervisor. All estimates include supervisor, month, and year fixed effects, and control for officer birth year, tenure, district, lagged arrests, lagged complaints, lagged overall crime rate, lagged violent crime rate, and the share of black supervisees. Wings depict 95% confidence intervals using robust standard errors. $N = 171,094$.

Estimates and standard errors for the black-white difference and Hispanic-white difference are reported in Table 6.

Three quarters before their evaluation ($EQ = -3$), white officers have a 1.89% chance of being nominated by their white supervisors. The following quarter, this likelihood increases by about 1 percentage-point (53%), then 1.7 percentage-points (90%) the quarter before their evaluation, then 1.94 percentage-points (103%) in the evaluation quarter. All of these estimates are significant at the 1% level.

After their evaluation, white officers are still more likely to be nominated relative to $EQ = -3$, but not as likely as they were in the quarter of their evaluation. Specifically, white officers are 1.2 percentage-points more likely to be nominated in the quarter following their evaluation—about a 38% decrease from the previous quarter—and 0.96 percentage-

points more likely to be nominated in $EQ = 2$ relative to $EQ = -3$. These estimates are significant at the 1% level and 10% level. Although white officers are less likely to be nominated after their evaluation quarter, their chances of nomination are still higher than the initial likelihood in $EQ = -3$. There also appears to be a sustained effect from the information acquisition that persists for two quarters after the officer's evaluation.

The story for black officers is a different one. Although the point estimates for black officers are positive leading up to their evaluation, they are not statistically significant until the quarter of their evaluation. In the evaluation quarter ($EQ = 0$), black officers are 0.57 percentage-points (61%) more likely to be nominated relative to the reference quarter, and this is significant at the 5% level. Immediately after their evaluation, the point estimates turn negative and are not statistically significant. Three quarters after their evaluation, the point estimate is negative (-0.0065) and significant at the 10% level. As the average black officer has a 0.94 chance of being nominated in $EQ = -3$, this means that their chances of nomination have been reduced by 69% in $EQ = 3$. Contrary to their white colleagues, black officers do not benefit from information acquisition after their evaluation.

Hispanic officers experience a statistically significant relative increase in nomination likelihood in the quarters before and of their evaluation: 1.2 percentage-points (67%) and 1.5 percentage-points (83%), respectively.³¹ After their evaluation, the point estimates are positive but not statistically significant. Like their black colleagues, the nomination likelihood three quarters after the evaluation for Hispanic officers is negative (-0.0086) and significant at the 10% level.

To summarize, white and Hispanic officers are more likely to be nominated in the quarters leading up to and including their evaluation. This boost grows steadily as we move closer to the evaluation quarter, and then drops after the evaluation quarter. In contrast, black officers assigned to white supervisors see a jump in their nomination likelihood only in the quarter of evaluation. Further, the negative black-white gap in all of the quarters are statistically

³¹The mean nomination probability for Hispanic officers assigned to white supervisors in $EQ = -3$ is 1.8 percent.

significant at either the 5% or 1% levels (Table 6).

Table 6: Racial Difference in Nomination Likelihood by Quarter

| Estimates for: | Outcome Variable: Nominated | | |
|--|-----------------------------|-------------------------|---------------------------|
| | White Officers (1) | Black-White Gap (2) | Hispanic-White Gap (3) |
| Quarter relative to three quarters before evaluation | | | |
| Two quarters pre-eval. (EQ = -2) | 0.0103*** (0.00277) | -0.00871** (0.00372) | -0.00659 (0.00472) |
| One quarter pre-eval. (EQ = -1) | 0.0170*** (0.00276) | -0.0130*** (0.00360) | -0.00521 (0.00454) |
| Evaluation quarter (EQ = 0) | 0.0194*** (0.00281) | -0.0139*** (0.00356) | -0.00468 (0.00444) |
| One quarter post-eval. (EQ = 1) | 0.0121*** (0.00296) | -0.0141*** (0.00354) | -0.00846* (0.00443) |
| Two quarters post-eval. (EQ = 2) | 0.00956*** (0.00318) | -0.0130*** (0.00378) | -0.00742 (0.00473) |
| Three quarters post-eval. (EQ = 3) | 0.00504 (0.00365) | -0.0116*** (0.00440) | -0.0136** (0.00536) |
| Observations | 171,094 | | |

Source: CPD analysis sample.

Notes: This table reports estimates of how the nomination likelihood changes by quarter. Estimates for white officers are in column 1, the black-white difference in column 2, and the Hispanic-white difference in column 3. See notes in Figure 1.

6 Experimental Evidence

Because our observational data do not capture interactions between supervisor and officer, we ran an online experiment to measure the review process in the nomination decision.³² Participants were asked to review CPD officer profiles and nominate one for an award. We study how officer race affects two types of choices: attention to an officer profile and the nomination decision. First, we measured which profiles participants hovered over, the order in which participants hovered over the profiles, and how long participants hovered over each profile. Second, we measured which officer was ultimately nominated for an award.

By using the same officers from the CPD analysis sample, we are able to generalize

³²The experiment was pre-registered in the AEA RCT Registry, AEARCTR-0005929.

our findings to a broader evaluator group than Chicago police supervisors. At the same time, we do not necessarily expect the two evaluator groups to act very differently; although Chicago police supervisors may be a selected sample, demographically-speaking, Dickinson et al. (2015) finds that police commissioners are no different from non-police civilians when it comes to issuing rewards.³³

6.1 Experimental Design

Survey participants were given two different types of tasks. In the first type of task, participants chose between a black male officer and a non-black male officer, where the black male officer was randomly assigned to be either “high-quality” or “low-quality” and the non-black male officer was assigned the converse.³⁴ In judging officer profiles, we used the number of civilian complaints and arrests. These classifications are admittedly subjective but they were made independently of officer race and sex. “High-quality” profiles were those with zero civilian complaints and an above-average number of arrests. “Low-quality” profiles were those with one or two civilian complaints and a below-average number of arrests.

In the second type of task, participants were shown four officer profiles and asked to nominate one for an award. In this task, officer profiles displayed only demographic information (e.g., race, sex, and age) and participants had to mouse over a profile to reveal full information about the officer.³⁵ All officers were of “average quality”, defined as having zero or one civilian complaints and an average number of arrests. There were two iterations of this task. In the first iteration, the officer group was racially heterogeneous. Three of the four profiles always featured a white officer, a black officer, and an Hispanic officer. The race of the fourth profile was randomly chosen amongst these three races. In the second iteration, three of the four profiles were always white officers and the race of the fourth profile was

³³In an experiment, Dickinson et al. (2015) finds that police commissioners are slightly more likely than non-police subjects to issue rewards but with less intensity. However, these differences are not statistically significant.

³⁴See Appendix Figure B2 for a screenshot of the task.

³⁵See Appendix Figure B3 for a screenshot of the task.

randomly chosen amongst white, black, and Hispanic.

For both types of tasks, the display ordering of officer profiles was randomly determined. All tasks were time-constrained to introduce a cost to reviewing profiles. Participants had 20 seconds to complete the first task (pairwise comparison) and 40 seconds to complete the second task (group comparison).³⁶ For the second type of task, participants were restricted from uncovering any work performance measures and from moving onto the next page for ten seconds. This was to ensure that participants had enough time to view and review the demographic information (e.g., race) of the four officer profiles on the screen. Although participants were asked to nominate an officer, they were not required to do so. That is, they were able to move onto the next page without nominating an officer. See Appendix B for more information about the online experiment.

6.2 Sample Selection and Data

The experiment was conducted on Amazon Mechanical Turk in July 2020.³⁷ We recruited 411 MTurk workers (hereafter “workers”) who were 18 years of age or older, based in the United States with English language proficiency, and who had access to a computer with a mouse and Javascript. The technical requirements were necessary in order to capture mouse movements on the screen. The survey had three data quality checks to identify bots and to ensure workers paid attention during the survey. For the analysis, we decided to include workers who passed at least two of the three data quality checks. This restriction reduces our final analysis sample to 407 workers.

³⁶These time limits appear to be within reason; participants took about 9.8 seconds, on average, for the pairwise comparison and 27.9 seconds, on average, for the group comparison. For the group comparison, conditional on mousing over any profile, about 70 percent of participants moused over all four profiles.

³⁷It is possible that the George Floyd incident on May 25, 2020 and subsequent protests may have altered people’s perceptions of the police and black individuals. Specifically, the incident may have increased MTurk workers’ interest in and affinity towards black officers because they are black. This would work against our results, which find that black officers are less likely to be moused over and are less likely to be nominated when paired against a non-black officer.

6.3 Results

Are black officers less likely to be nominated for an award?

Columns 1 through 3 of Table 7 reports results from the pairwise comparison of a black and non-black (white or Hispanic) officer. Column 1 reports results from all MTurk workers, column 2 is restricted to white MTurk workers, and column 3 is restricted to black MTurk workers. Low-quality black officers are 8.8 percentage-points ($p < 0.01$) less likely than low-quality white officers to be nominated. This gap largely persists with high-quality officers. Although high-quality officers are almost 50 percentage-points more likely to be nominated for an award ($p < 0.01$) than low-quality officers, high-quality black officers are still 5.6 percentage-points less likely to be nominated than high-quality white officers. This difference is statistically significant at the 10 percent level. When we focus on the race of the survey participants, we see that the results are driven by white workers (column 2). The black-white gap among white workers is -10 percentage points ($p < 0.01$) for low-quality officer profiles and -5.5 percentage points for high-quality officer profiles (p-value = 0.111). There is no statistically significant black-white nomination gap among black workers. These results are consistent with the CPD analysis, which found a negative black-white nomination gap among white supervisors but not among black supervisors.

Table 7: Impact of being Black on Nomination Likelihood

| Pairwise Comparison: Race of MTurk Worker: | Outcome Variable: Nominated | | | |
|---|-----------------------------|-----------------------|---------------------|---------------------|
| | Black v. Non-Black | | | High v. Low |
| | All (1) | White (2) | Black (3) | All (4) |
| High-Quality Profile | 0.483*** (0.0921) | 0.410*** (0.0947) | 0.771*** (0.284) | 0.524*** (0.169) |
| Black Officer | -0.0883*** (0.0311) | -0.103*** (0.0361) | -0.0163 (0.0823) | -0.0324 (0.119) |
| High-Quality x Black Officer | 0.0326 (0.0448) | 0.0487 (0.0518) | -0.00390 (0.121) | 0.141 (0.171) |
| Female Officer | | | | -0.0705 (0.0680) |
| High-Quality x Female Officer | | | | 0.137 (0.107) |
| Observations | 1,576 | 1,196 | 256 | 794 |

Source: MTurk survey data.

Notes: This table reports estimates from a pairwise comparison of officer profiles. Columns 1-3 are a pairwise comparison between a black male officer and a non-black male officer. Column 4 is a pairwise comparison between two officers of the same race and sex but differing profile qualities. All estimates control for officer traits and profile location on screen. Officer traits include officer age, tenure, arrests, and complaints. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In column 4, we conduct a robustness check wherein the two officer profiles are of the same race and sex and differ only in terms of quality. As expected, high-quality profiles are more likely to be nominated—about 52 percentages-points—and this is significant at the 1 percent level. This also provides an indirect test that MTurk workers were able to discern the quality difference between the two officer profiles. Interestingly, when the officers are both black or both female, workers are even more likely to nominate the high-quality officer (about 14 additional percentage-points for black and female officers) relative to when the officers are white males. Although these estimates are not statistically significant, the positive point estimates, together with the results from column 1, suggest that either white males are given some slack even if they do not meet a certain standard or that minorities are held to a higher standard.

Do MTurk workers choose different levels of attention based on officer’s race?

Table 8 presents summary statistics on information acquisition by MTurk workers for white officers, black officers, and Hispanic officers. We examine three different measures of information acquisition: ever moused over (Panel A), first mouseover (Panel B), and mouseover duration measured in seconds (Panel C). Looking at all officers, the first row in Panel A tells us that workers tend to mouse over most of the officer profiles: over 80 percent of officer profiles were moused over. Specifically by race, 84.2 percent of white officer profiles were moused over, 81.5 percent of black officer profiles were moused over, and 81.8 percent of Hispanic officer profiles were moused over. The white-black difference is borderline significant, with a p-value of 0.107. The white-Hispanic difference is not statistically significant.

When we separate the results by the racial composition of the group, however, we see a wider gap between white and black officers. Black officer profiles are 7.3 percentage points less likely to be moused over compared to white officer profiles, and this difference is significant at the 5 percent level. This means that when the minority officer in a predominantly white officer pool is black, the black officer is less likely to be moused over. However, if the minority officer is Hispanic, then there is no difference in mouse-over likelihood.

Conditional on being moused over, there does not appear to be a significant black-white difference regarding which officer is moused over first (Panel B). However, there is a difference in the amount of time spent reviewing profiles. Workers spend around half a second more reviewing black and Hispanic profiles, and these are significant at the 1 percent level. Table 9 reports the findings in a regression framework, where we control for the profile location on the computer screen and the worker’s starting mouse position.

Taken together, these results appear to suggest that there is a dichotomy in the type of workers who are likely to mouse over minority profiles. Some workers are less interested in acquiring information about black officers. Among workers who are interested, however, black and white officers are equally likely to be the first mouseover.

Table 8: Information Acquisition by Officer Race, Comparison of Means

| | White Officer | Black Officer | Hispanic Officer | B-W Difference (p-value) | H-W Difference (p-value) |
|---|------------------|------------------|---------------------|-----------------------------|-----------------------------|
| Panel A: Outcome Variable: Ever Moused Over | | | | | |
| All Officers | 84.2% | 81.5% | 81.8% | -0.028 (0.107) | -0.024 (0.163) |
| Het. Race Officer Group | 85.2% | 82.7% | 80.6% | -0.025 (0.276) | -0.046 (0.054) |
| Predom. White Officer Group | 83.8% | 76.6% | 86.7% | -0.073 (0.036) | 0.028 (0.420) |
| Panel B: Outcome Variable: First Mouseover | | | | | |
| All Officers | 30.0% | 31.9% | 28.4% | 0.019 (0.419) | -0.016 (0.490) |
| Het. Race Officer Group | 32.8% | 31.0% | 26.5% | -0.018 (0.581) | -0.063 (0.048) |
| Predom. White Officer Group | 28.9% | 35.7% | 35.6% | 0.069 (0.155) | 0.067 (0.152) |
| Panel C: Outcome Variable: Mouseover Duration (seconds) | | | | | |
| All Officers | 2.33 | 2.78 | 2.89 | 0.448 (0.000) | 0.559 (0.000) |
| Het. Race Officer Group | 2.27 | 2.64 | 2.85 | 0.366 (0.006) | 0.576 (0.000) |
| Predom. White Officer Group | 2.36 | 3.40 | 3.07 | 1.041 (0.000) | 0.714 (0.001) |

Source: MTurk survey data.

Notes: This table reports mean values for the three measures of information acquisitions: ever moused over in Panel A, first mouseover in Panel B, and mouseover duration in Panel C. B-W Difference reports the percentage-point difference between black officers and white officers. H-W Difference reports the percentage-point difference between Hispanic officers and white officers. p-values are the p-value from a t-test of a difference in means.

One interesting fact is that, conditional on mousing over a profile, workers spend about half a second longer on minority profiles (Panel C). Further, Hispanic officer profiles are less likely to be moused over, vis-a-vis white officer profiles, when the officer pool is more racially heterogeneous. In that case, the Hispanic-white difference is 4.6 percentage points and this difference is significant at the 10 percent level. Conditional on being moused over, Hispanic officers are also less 6.3 percentage points less likely to be the first mouseover compared to white officers. This difference is significant at the 5 percent level. Again, this difference appears only when the officer pool is racially heterogeneous. One potential interpretation of these results is that when the two racial minorities are in the same group (black officers and Hispanic officers), they compete with each other for attention and black officers edge out Hispanic officers. We are uncertain of why this may be the case, but it is possible that the George Floyd protests, which took place about a month prior to the online experiment, may have affected workers' decisions on who to mouseover first.

Table 9: Impact of Officer Race on Information Acquisition

| Officer Pool: | All (1) | Het. Race (2) | Predom. White (3) |
|---|---------------------|-----------------------|----------------------|
| Panel A: Outcome Variable: Ever Moused Over | | | |
| Black Officer | -0.0289 (0.0177) | -0.0271 (0.0230) | -0.0707* (0.0386) |
| Hispanic Officer | -0.0252 (0.0179) | -0.0475** (0.0237) | 0.0332 (0.0329) |
| Observations | 2,992 | 1,500 | 1,492 |
| Mean Outcome for White Officer | 0.842 | 0.852 | 0.838 |
| Panel B: Outcome Variable: First Mouseover | | | |
| Black Officer | 0.00355 (0.0210) | -0.0285 (0.0293) | 0.0399 (0.0409) |
| Hispanic Officer | -0.0136 (0.0212) | -0.0578** (0.0293) | 0.0681 (0.0424) |
| Observations | 2,488 | 1,243 | 1,245 |
| Mean Outcome for White Officer | 0.300 | 0.328 | 0.289 |
| Panel C: Outcome Variable: Mouseover Duration (seconds) | | | |
| Black Officer | 0.431*** (0.107) | 0.347** (0.135) | 0.984*** (0.230) |
| Hispanic Officer | 0.570*** (0.107) | 0.575*** (0.137) | 0.697*** (0.224) |
| Observations | 2,488 | 1,243 | 1,245 |
| Mean Outcome for White Officer | 2.335 | 2.270 | 2.361 |

Source: MTurk survey data.

Notes: This table reports estimates for racial differences in a group comparison of officer profiles. We examine three different measures of information acquisition: ever moused over in Panel A, first mouseover in Panel B, and mouseover duration in Panel C. All estimates control for profile location on screen and worker's starting mouse position. Standard errors are in parentheses. Panels A and B report robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Next, we examine how information acquisition affects the probability of nomination. Table 10 reports the probability of nomination conditional on officer race and the information-acquisition measure: ever moused over (Panel A), first mouseover (Panel B), and mouseover duration (Panel C).

Panel A finds that profiles that were moused over are 8.8 percentage points more likely to be nominated, on average. This indicates that mouseover activity is a good measure of

worker interest in the officer. Workers who mouse over black or Hispanic officers are even more likely to nominate these minority officers, by an additional 8.7 percentage points and 14 percentage points, respectively. Looking at column 3, we see that these results are driven by the predominantly white officer pool.

Conditional on being moused over, being moused over first increases the probability of nomination by 2.5 percentage points though this is not statistically significant (Panel B). Relative to white officers, however, black officers who are moused over first are 9.2 percentage-points even more likely to be nominated. This estimate is significant at the 10 percent level. Column 3 indicates that this finding is driven by a more racially homogeneous officer pool. Since a minority officer will stand out in a predominantly white officer pool, if a worker chooses to mouse over the minority officer first, then the worker is also decidedly more likely to nominate the minority officer (19 percentage points for black officers and 17 percentage points for Hispanic officers). As the mean nomination rate for white officers is 21.6 percent, these estimates mean that being moused over first increases the probability of nomination by 89 percent for black officers and 79.6 percent for Hispanic officers in comparison to white officers.

The results in Panels A and B are consistent with the narrative that there is a dichotomy among workers who are likely to mouse over minority profiles vs. not. Some workers are more interested in acquiring information about black and Hispanic officers, and those workers are also more likely to nominate minority officers. Because all officer profiles were of average quality, the decision to nominate an officer should be independent of officer performance and depend solely on the evaluator's preferences. (All estimates control for the officer profile's location and the worker's starting mouse position on the screen.)

Panel C looks at the unconditional impact of mouseover duration on nomination likelihood.³⁸ There is a baseline black-white difference in nomination likelihood of -0.019, which is not statistically significant. The longer a worker spends viewing an officer's profile, the

³⁸In other words, we include all officers profiles in the analysis, even the ones that were not moused over.

higher the chance of a nomination: an increase of 4.4 percentage points for each additional second. This is significant at the 1 percent level. This estimate does not differ for black officers; each additional second on a white officer's profile increases the probability of nomination the same as an additional second spent on a black officer's profile. However, Hispanic officers do benefit; each additional second increases the nomination likelihood by an additional 2.6 percentage points relative to white officers. This estimate is significant at the 1 percent level.

Column 3 reports estimates when the officer pool is predominantly white—similar to the Chicago Police Department context. The baseline black-white difference in a predominantly white officer pool is -0.198 and significant at the 1 percent level. This is likely driven by workers who do not mouse over black officer profiles, as mouseover duration in that instance is zero seconds. The coefficient on the interaction between mouseover duration and black officers is 0.039, which means that an additional second spent reviewing a black profile is associated with an additional 3.9 percentage point increase in nomination likelihood relative to time spent reviewing a white profile. Given the large, negative baseline black-white nomination gap, however, this suggests that a worker would need to spend about five additional seconds on a black officer's profile in order to equalize the nomination probability for white vs. black officers. Interestingly, any Hispanic-white differences become insignificant in a predominantly white officer pool.

Table 10: Impact of Information Acquisition on Nomination Likelihood

| Officer Pool: | Outcome: Nominated Officer | | |
|---------------------------------------|----------------------------|-----------------------|------------------------|
| | All (1) | Het. Race (2) | Predom. White (3) |
| Panel A: Ever Moused Over | | | |
| Ever Moused Over | 0.0882*** (0.0244) | 0.107** (0.0495) | 0.0785*** (0.0284) |
| Black Officer | -0.0629 (0.0388) | -0.294 (0.251) | -0.380*** (0.0771) |
| Hispanic Officer | -0.157*** (0.0378) | -0.0177 (0.145) | -0.0102 (0.0684) |
| Ever Moused Over x Black Officer | 0.0865* (0.0443) | -0.00966 (0.0649) | 0.364*** (0.0797) |
| Ever Moused Over x Hispanic Officer | 0.140*** (0.0412) | 0.104* (0.0613) | 0.173** (0.0812) |
| Observations | 2,992 | 1,500 | 1,492 |
| Panel B: First Mouseover | | | |
| First Mouseover | 0.0251 (0.0259) | 0.109** (0.0483) | -0.00466 (0.0313) |
| Black Officer | -0.00694 (0.0268) | -0.205 (0.302) | -0.138* (0.0834) |
| Hispanic Officer | -0.0330 (0.0326) | 0.0977 (0.163) | 0.110* (0.0600) |
| First Mouseover x Black Officer | 0.0922* (0.0493) | -0.0102 (0.0669) | 0.194* (0.106) |
| First Mouseover x Hispanic Officer | 0.0242 (0.0528) | -0.113 (0.0701) | 0.172* (0.102) |
| Observations | 2,488 | 1,243 | 1,245 |
| Panel C: Mouseover Duration (seconds) | | | |
| Mouseover Duration | 0.0442*** (0.00638) | 0.0544*** (0.0116) | 0.0397*** (0.00763) |
| Black Officer | -0.0191 (0.0258) | -0.347 (0.229) | -0.198*** (0.0726) |
| Hispanic Officer | -0.105*** (0.0303) | 0.0550 (0.128) | 0.0863 (0.0582) |
| Mouseover Duration x Black Officer | 0.0102 (0.00966) | -0.0127 (0.0150) | 0.0387** (0.0173) |
| Mouseover Duration x Hispanic Officer | 0.0260*** (0.00929) | 0.0188 (0.0141) | 0.0140 (0.0150) |
| Observations | 2,992 | 1,500 | 1,492 |
| Mean Outcome for White Officer | 0.225 | 0.245 | 0.216 |

Source: MTurk survey data.

Notes: This table reports estimates for racial differences in the impact of information acquisition on nomination likelihood in a group comparison of officer profiles. We examine three different measures of information acquisition: ever moused over in Panel A, first mouseover in Panel B, and mouseover duration in Panel C. All estimates control for officer traits, profile location on screen, and worker's starting mouse position. Officer traits include officer age, tenure, arrests, and complaints. Standard errors are in parentheses. Panel A reports robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

To summarize the results, we find that black officers are less likely to be nominated compared to white officers regardless of profile quality. In terms of the information acquisition process, we find that black officers are less likely to be moused over, and this is more salient when workers are choosing among three white officers and one black officer. We do not see similar patterns when workers are choosing among three white officers and one Hispanic officer. Taken together, these findings suggest that the negative black-white nomination gap is due to a lack of interest about black officers on the part of workers when choosing between black and white officers.

We also find that, conditional on being moused over, black officers are as equally likely as white officers to be the first mouseover, but workers that choose to mouse over a black profile first are also more likely to nominate black officers over white officers. This suggests that black officers may benefit from having supervisors who are interested in interacting with them. Further, we also find suggestive evidence that black officers are under more scrutiny when they are the sole minority.

7 Discussion

In this section, we discuss whether the observed racial disparity is due to statistical discrimination or taste-based discrimination. The basis for statistical discrimination comes from the evaluator’s inability to observe or the lack of knowledge about a specific individual’s productivity. Because of this incomplete information, evaluators make choices based on the group’s mean performance. In our context, where supervisors are nominating officers for a specific incident, the negative black-white nomination gap conditional on arrest activity does not appear to be due to statistical discrimination. Now, one can argue whether arrests are the correct productivity measure. We would argue that it is not, but it is difficult to know whether officers agree without information on their beliefs. Further, even if officers would agree that arrests are not a true measure of productivity, the problem of being unable to

accurately measure productivity remains. For example, one may argue that an officer who breaks up a fight and de-escalates a situation is just as productive as an officer who arrests a suspect who committed a robbery. However, the officer in the first situation did not make any arrests. At the same time, irrespective of whether or not arrests are an accurate proxy for productivity, it does appear that arrests are rewarded in the police department.

Without officer beliefs, one method of discerning whether discrimination is due to lack of information, is to examine how the black-white nomination gap changes (or not) as more information about the officer is revealed (Altonji and Pierret, 2001). Figure 1 attempts this exercise, by looking at the black-white nomination gap by quarter relative to the annual evaluation. The fact that we still see a negative black-white nomination in the evaluation quarter suggests that taste-based discrimination may be at play.

There is some support for this theory in the MTurk analysis. We find that black officers are less likely to be moused over when in a predominantly white officer pool, but Hispanic officers do not have same experience (Table 9). Further, not being moused over indeed translates to worse outcomes for the officers; the black-white mouseover gap is -0.07 percentage points, but the black-white nomination gap conditional on not being moused over is -0.38.

Due to data limitations and the nature of supervisor assignments in the CPD, it is difficult to conclusively ascertain the type of discrimination. However, we believe that our paper leverages rich data and institutional knowledge to answer a first-order question, which is whether officers exhibit racial bias against fellow officers.

8 Application to Law Enforcement and Conclusion

Racial bias has been extensively documented in a variety of settings, including hiring decisions (Agan and Starr, 2017; Bertrand and Mullainathan, 2004; Craigie, 2020; Doleac and Hansen, 2018), sports umpires (Parsons et al., 2011), judicial and sentencing decisions

(Flanagan, 2018; Mueller-Smith and Schnepel, 2017; Park, 2017; Rehavi and Starr, 2014)³⁹, and bail decisions (Arnold et al., 2018). The increasing availability of police administrative data has allowed researchers to carefully examine and detect bias in law enforcement as well.

A potential solution that has been put forth is to increase racial and gender diversity among officers, who are traditionally homogeneous.⁴⁰ A diverse police force may improve policing quality in various ways (Sklansky, 2005). Outwardly, it may improve the police’s relationship with the community through unique skills that minority officers may possess (Anwar et al., 2012; Harvey and Mattia, 2020; Miller and Segal, 2018).⁴¹ Inwardly, it may alter the internal dynamics of the department.

This paper asks how racial bias affects career progression, which is of particular relevance to law enforcement, where minorities are less represented at higher ranks. For example, white males comprised 40 percent of all entry-level police officers in the Chicago Police Department in 2015, and 56 percent of those at the rank of Sergeant or higher.⁴² In addition to improving policing quality⁴³, minority representation—particularly at higher ranks of office—may help to recruit more minorities and close promotion gaps, which may further attract minority applicants. Prior research has found that minorities in management positions can address wage gaps and occupational gaps (Kofoed and McGovney, 2019; Langan, 2018). At the same time, an extensive literature documents racial and gender bias in the workplace, which may

³⁹Mueller-Smith and Schnepel (2017) finds that the practice of diversion, or a halt or termination of one’s progression through the justice system, reduces re-offending rates and improves labor market outcomes among young black men charged with misdemeanors.

⁴⁰For example, in their investigative report of the Ferguson Police Department, the U.S. Department of Justice called for a more diverse police force as part of a broader reform effort (U.S. Department of Justice, 2015, p. 58). Several cities, including Chicago, Indianapolis, and Knoxville, have followed this lead and pursued diversity initiatives (Chicago: <https://www.chicagotribune.com/news/breaking/ct-met-chicago-police-hiring-20180503-story.html>; Indianapolis: <https://www.indystar.com/story/opinion/columnists/suzette-hackney/2018/09/27/impd-leads-charge-toward-diversity-columnist-suzette-hackney-writes/1433649002/>; Knoxville: <https://www.knoxnews.com/story/news/local/2017/01/09/knoxville-police-department-recruits-remain-diverse-group/96345092/>)

⁴¹McCrary (2007) and Garner et al. (2020) do not find that court-ordered affirmative action litigation affects offense and arrest rates, but Garner et al. (2020) acknowledges that there may be racially heterogeneous effects that offset each other.

⁴²These numbers do not include civilian Department members.

⁴³See, for example, Bulman (2019); Donohue III and Levitt (2001); Miller and Segal (2018).

hinder minorities’ career progression.⁴⁴ In the context of policing, diversity initiatives may be constrained by the extent to which officer bias carries over to their colleagues. Indeed, in the Chicago Police Department, 98 percent of CPD officers believe that promotions are due to connections not merit (Police Accountability Task Force, 2016).

To examine the extent of racial bias in law enforcement, we construct a panel dataset of all CPD officers containing their personnel information. We exploit quasi-random variation in supervisor assignment and find that white supervisors are less likely to nominate black officers than white or Hispanic officers. These results control for officer characteristics, including experience, arrest record, and misconduct. We find that supervisors are more likely to nominate officers in the evaluation quarter, suggesting that interactions and information acquisition is important for career recognition, but also that black officers benefit less than their white peers.

To supplement our CPD analysis, we run an online experiment using Amazon Mechanical Turk workers and, again, find that black officers are less likely to be nominated than their non-black peers. In terms of the information acquisition process, we find that black officers are less likely to be moused over and, conditional on being moused over, are reviewed for longer.

Our findings have two important policy implications for law enforcement. First, we find that interactions between supervisors and officers are an important mechanism for career recognition, suggesting that the observed racial gap in award nominations may be due to lack of information. One way to overcome discrimination that is driven by biased beliefs is continuous and sustained positive evaluations (Bohren et al., 2019). But, as we find a persistent negative black-white gap even in the evaluation quarter, our results suggest that the decentralized nature of supervision and oft-changing supervisor assignment in the CPD

⁴⁴For example, Egan et al. (2018) find that female financial advisors are 20% more likely than male financial advisors to lose their jobs following a misconduct. In medicine, Sarsons (2019) finds that physicians are less likely to refer to new female surgeons after a bad patient outcome but not to new male surgeons. Beaman et al. (2018) find that women are harmed in a referral-based hiring process as workplace networks tend to be gender homophilous. Glover et al. (2017) find that manager bias can cause a self-fulfilling prophecy in that biased managers interact less with minority cashiers, leading them to exert less effort.

present a challenge for discrimination to be reversed.

Second, our finding of a persistent black-white recognition gap suggests that simply increasing the diversity of incoming recruits may not be enough to eliminate racial bias in policing. The argument for a diverse police force stems from the “contact hypothesis”, or that outsider bias can be reduced if the integrated group has a common goal. Although there is empirical evidence in support of this theory (Lowe, 2019), another study finds that the improved behavior towards out-group members does not extend beyond the intervention setting (Mousa, 2020). Therefore, it is uncertain whether focusing on the diversity of incoming officers will be enough to eliminate racial bias in the department. Further, biased evaluations may lead the discriminated party to exert less effort and have lower performance, affecting pay and promotions (MacLeod, 2003). As such, police departments should also pursue policies that address in-group bias due to its effect on career advancement.

References

- Agan, A. and S. Starr (2017). Ban the box, criminal records, and racial discrimination: A field experiment. *The Quarterly Journal of Economics* 133(1), 191–235.
- Ajilore, O. and S. Shirey (2017). Do #AllLivesMatter? An evaluation of race and excessive use of force by police. *Atlantic Economic Journal* 45(2), 201–212.
- Alsan, M., O. Garrick, and G. Graziani (2019). Does diversity matter for health? experimental evidence from oakland. *American Economic Review* 109(12), 4071–4111.
- Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The quarterly journal of economics* 116(1), 313–350.
- Antonovics, K. and B. G. Knight (2009). A new look at racial profiling: Evidence from the boston police department. *The Review of Economics and Statistics* 91(1), 163–177.
- Anwar, S., P. Bayer, and R. Hjalmarsson (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics* 127(2), 1017–1055.
- Anwar, S. and H. Fang (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review* 96(1), 127–151.
- Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics* 133(4), 1885–1932.
- Ba, B., D. Knox, J. Mummolo, and R. Rivera (2020). Diversity in Policing: The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago. Working paper.
- Bacher-Hicks, A. and E. de la Campa (2020). Social Costs of Proactive Policing: The Impact of NYC’s Stop and Frisk Program on Educational Attainment. Working paper.
- Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review* 106(6), 1437–75.
- Beaman, L., N. Keleher, and J. Magruder (2018). Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi. *Journal of Labor Economics* 36(1), 121–157.
- Bertrand, M. and S. Mullainathan (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.
- Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American economic review* 109(10), 3395–3436.
- Bulman, G. (2019). Law Enforcement Leaders and the Racial Composition of Arrests. *Economic Inquiry*.

- Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics* 125(3), 1101–1144.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9), 2593–2632.
- Close, B. R. and P. L. Mason (2006). After the traffic stops: Officer characteristics and enforcement actions. *The BE Journal of Economic Analysis & Policy* 6(1).
- Craigie, T.-A. (2020). Ban the box, convictions, and public employment. *Economic Inquiry* 58(1), 425–445.
- Cullen, Z. B. and R. Perez-Truglia (2020). The old boys’ club: Schmoozing and the gender gap. Working paper, National Bureau of Economic Research.
- Cunningham, J. P. and R. Gillezeau (2018, May). Racial differences in police use of force: Evidence from the 1960s civil disturbances. *AEA Papers and Proceedings* 108, 217–21.
- Dickinson, D. L., D. Masclet, and M. C. Villeval (2015). Norm enforcement in social dilemmas: An experiment with police commissioners. *Journal of Public Economics* 126, 74–85.
- Doleac, J. L. and B. Hansen (2018). Does ”ban the box” help or hurt low-skilled workers? statistical discrimination and employment outcomes when criminal histories are hidden. *Journal of Labor Economics* Forthcoming.
- Donohue III, J. J. and S. D. Levitt (2001). The impact of race on policing and arrests. *The Journal of Law and Economics* 44(2), 367–394.
- Egan, M. L., G. Matvos, and A. Seru (2018). When Harry fired Sally: The Double Standard in Punishing Misconduct. Working paper, National Bureau of Economic Research.
- Flanagan, F. X. (2018). Race, gender, and juries: Evidence from north carolina. *The Journal of Law and Economics* 61(2), 189–214.
- Garner, M., A. Harvey, and H. Johnson (2020). Estimating effects of affirmative action in policing: A replication and extension. *International Review of Law and Economics* 62, 105881.
- Gershenson, S., C. Hart, J. Hyman, C. Lindsay, and N. W. Papageorge (2018). The long-run impacts of same-race teachers. Working paper, National Bureau of Economic Research.
- Giuliano, L., D. I. Levine, and J. Leonard (2009). Manager race and the race of new hires. *Journal of Labor Economics* 27(4), 589–631.
- Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics* 132(3), 1219–1260.

- Goncalves, F. and S. Mello (2020). A Few Bad Apples?: Racial Bias in Policing. Working paper.
- Harvey, A. and T. Mattia (2020). Reducing racial disparities in crime victimization. Technical report.
- Hengel, E. (2019). Publishing while female. are women held to higher standards? evidence from peer review.
- Hoekstra, M. and C. Sloan (2020). Does Race Matter for Police Use of Force? Evidence from 911 Calls. Working paper.
- Horrace, W. C. and S. M. Rohlin (2016). How dark is dark? bright lights, big city, racial profiling. *Review of Economics and Statistics* 98(2), 226–232.
- Kessler, J. B., C. Low, and C. D. Sullivan (2019). Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review* 109(11), 3713–44.
- Knowles, J., N. Persico, and P. Todd (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy* 109(1), 203–229.
- Knox, D., W. Lowe, and J. Mummolo (2020). Administrative Records Mask Racially Biased Policing. *American Political Science Review* 114(3), 619–637.
- Kofoed, M. and E. McGovney (2019). The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at west point. *Journal of Human Resources* 54(2), 430–467.
- Lang, K. and A. Kahn-Lang Spitzer (2020). Race discrimination: An economic perspective. *Journal of Economic Perspectives* 34(2), 68–89.
- Langan, A. (2018). Female Managers and Gender Disparities: The Case of Academic Department Chairs. Working paper.
- Lowe, M. (2019). Types of contact: A field experiment on collaborative and adversarial caste integration. Working paper.
- MacLeod, W. B. (2003). Optimal contracting with subjective evaluation. *American Economic Review* 93(1), 216–240.
- Mason, P. L. (2007). Driving while black: do police pass the test? *Swedish Economic Policy Review* 14, 79–113.
- McCrary, J. (2007). The Effect of Court-ordered Hiring Quotas on the Composition and Quality of Police. *American Economic Review* 97(1), 318–353.
- Miller, A. and C. Segal (2018). Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence. *Review of Economic Studies* Accepted.

- Mousa, S. (2020). Building social cohesion between christians and muslims through soccer in post-isis iraq. *Science* 369(6505), 866–870.
- Mueller-Smith, M. and K. Schnepel (2017). Diversion in the criminal justice system: Regression discontinuity evidence on court deferrals. Technical report, Working paper.
- Neumark, D., I. Burn, and P. Button (2019). Is it harder for older workers to find jobs? new and improved evidence from a field experiment. *Journal of Political Economy* 127(2), 922–970.
- Nix, J., B. A. Campbell, E. H. Byers, and G. P. Alpert (2017). A Bird’s Eye View of Civilians Killed by Police in 2015. *Criminology and Public Policy* 16(1), 309–340.
- Owens, E., D. Weisburd, K. L. Amendola, and G. P. Alpert (2018). Can you build a better cop? experimental evidence on supervision, training, and policing in the community. *Criminology & Public Policy* 17(1), 41–87.
- Park, K. H. (2017). The impact of judicial elections in the sentencing of black crime. *Journal of Human Resources* 52(4), 998–1031.
- Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review* 101(4), 1410–35.
- Police Accountability Task Force (April 2016). Recommendations for reform: Restoring trust between the chicago police and the communities they serve. Technical report.
- Rehavi, M. M. and S. B. Starr (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy* 122(6), 1320–1354.
- Rim, N., B. Ba, and R. Rivera (2020). Disparities in Police Award Nominations: Evidence from Chicago. *AEA Paper and Proceedings* 110, 447–451.
- Sarsons, H. (2019). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. Working paper.
- Sarsons, H. (2020). Gender differences in recognition for group work. *Journal of Political Economy Accepted*.
- Sklansky, D. A. (2005). Not Your Father’s Police Department: Making sense of the new demographics of law enforcement. *J. Crim. L. & Criminology* 96(3), 1209–1243.
- U.S. Department of Justice (January 13, 2017). Investigation of the Chicago Police Department. Technical report.
- U.S. Department of Justice (March 4, 2015). Investigation of the Ferguson Police Department. Technical report, Civil Rights Office.
- Weisburd, E. (2018). Whose Help Is on the Way? The Importance of Individual Police Officers in Law Enforcement. Working paper.

- West, J. (2018). Racial Bias in Police Investigations. Working paper.
- Worden, R. E., M. Kim, C. J. Harris, M. A. Pratte, S. E. Dorn, and S. S. Hyland (2013). Intervention with problem officers: An outcome evaluation of an eis intervention. *Criminal justice and behavior* 40(4), 409–437.
- Zeltzer, D. (2020). Gender homophily in referral networks: Consequences for the medicare physician earnings gap. *American Economic Journal: Applied Economics* 12(2), 169–97.