The Black-White Recognition Gap in Award Nominations*

Nayoung Rim† Roman Rivera‡ Andrea Kiss§ Bocar Ba¶

April 4, 2021

There is substantial evidence showing racial bias in firms' hiring decisions, but less is known about bias in career recognition. We construct a novel dataset of police award nominations to measure bias against minority employees. Exploiting quasi-random variation in supervisor assignment and randomized timing of annual evaluations, we find that white supervisors are less likely to nominate black officers than white and Hispanic officers leading up to and during the evaluation period. Further, the black-white nomination gap widens with the number of arrests. These patterns suggest that the disparity is not due to in-group favoritism towards white officers but rather bias against black officers. We conduct an online experiment to examine evaluator engagement and find that evaluators are less likely to engage with black officers vs. white officers. Our findings suggest bias in career recognition may have important implications for the black-white promotion gap, the lack of diversity in upper-management positions, and, ultimately, the racial wage gap.

JEL Codes: J71, M51, J45, J48
Keywords: racial bias, award nominations, promotions, supervisors, police

*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 necessarily 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 engage 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. If the cost of acquiring information differs by race, a racial gap in career recognition and progression may result. 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 due to a union contract, and promotions are largely determined by a written

[1] In 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.

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, which we use to approximate random assignment of an officer's race to a supervisor. We confirm as-good-as-random assignment by analyzing supervisor-officer assignments and confirm 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 engage with and gather information about officers in 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.

We find that supervisors are more likely to nominate all officers in the quarter of evaluation relative to two quarters prior, suggesting that statistical discrimination may exist (Altonji and Pierret, 2001).[5] However, black officers are nominated 35.4 to 42.7 percent less than white officers, suggesting that statistical discrimination is not the only explanation for this racial disparity. The fact that the nomination likelihood increases for all officers in the evaluation quarter suggests that white supervisors may be learning about their officers, but the persistent, negative black-white disparity suggests that the learning

[2]Seventy 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.

[3]In 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 nomination gap among white supervisors even in the absence of racial bias.

[4]The 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.

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

is not manifesting in changed behavior towards black officers. By contrast, nomination patterns for Hispanic officers are similar to those for white officers. These results suggest that the black-white gap is not due to in-group favoritism towards white officers but rather bias against black officers. This is supported by additional analysis that finds that the negative black-white gap among white supervisors widens with the number of arrests. This is the opposite of what we would expect if the disparity were due to white supervisors' inaccurate beliefs about black officers' work measures.

Because the administrative 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 within a time limit. In addition to being able to experimentally examine whether evaluators choose to engage with minority officers, the online experiment allows us to generalize our findings to a broader evaluator group than Chicago police supervisors.

In one task, evaluators choose between a black officer and a non-black officer. Although officer performance levels are randomly chosen, evaluators 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 evaluators must mouse over the profile to reveal full information about the given officer. We monitor mouse movements across the screen and find that black profiles are least likely to be moused over. The black-white difference in engagement more than doubles (from -2.8 percentage points to -7.3 percentages points) when evaluators are choosing among three white officers and one black officer. We do not see similar patterns for Hispanic officers when evaluators are choosing among three white officers and one Hispanic officer. When the officer pool becomes more racially diverse, any benefits black officers had received from greater evaluator engagement disappear. Taken together, our findings suggest that racial issues in policing are not just at issue between police and the public, but also within departments, and thus that simply hiring minority

officers may be limited in its efficacy.

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-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 evaluators, both groups that are mostly white, are less likely to gather information on minority officers, leading to a racial disparity in award nominations. 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 evaluators are less likely to engage with black officers, particularly when black officers are in a pool with three white officers. Additionally, evaluators spend more time evaluating black 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-

[6]Hengel (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). Recent papers have sought to use officer dispatches to 911 calls or to investigate automobile crashes to address this issue (Hoekstra and Sloan, 2020; Weisburst, 2018; West, 2018). Our paper attempts to bypass the truncated data problem by focusing on supervisor nominations of as-good-as-randomly assigned officers. Importantly, 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 discusses the identifying assumptions. We present results using administrative CPD data in Section 5 and the experimental evidence in Section 6. We conclude with a discussion of the policy implications for law enforcement agencies in Section 7.

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 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

[7] See, 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); Weisburst (2018); West (2018); **?**.

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

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 (e.g., protecting the scene, establishing the perimeter), 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 on 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 advancement, such as performance evaluations[12], merit promotions[13], and overtime

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

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

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

[12]Chicago Police Department, Career Development Directive, Employee Resource E05-01, Section IV.H., available at http://directives.chicagopolice.org.

[13]Merit promotions are a more subjective selection process that relies on a variety of officer qualifications vis-a-vis test-based promotions. See, Section III.E.2, Employee Resource E05-05, available at http://directives.chicagopolice.org.

pay[14].

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 2016 report by the Police Accountability Task Force, 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. For example, 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 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 quasi-random assignment and the annual evaluation requirement.

[14]The 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 estimate for an additional arrest last year ($46.99). These estimates control for years of experience.

[15]We 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.

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

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 the 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),

[17] There 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, available at http://directives.chicagopolice.org).

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

covering 2002 to 2017.

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 geographic patrol 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 to 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] Crimes are aggregated into three categories: violent crime, property crime, and non-index crime. The Federal Bureau of Investigation classifies violent and property crimes as "index crimes" because

[19] An 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.

[20] Most 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.

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

they are more serious offenses.[22] Non-index crimes capture crimes that are not related to violence or property, such as municipal code violations, traffic violations, warrants, drugs, prostitution, and gambling.[23]

Complaints The complaints dataset contains 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 to 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

[22] Violent crimes are crimes related to violence, such as murder and assault. Property crimes are crimes related to property, such as burglary and motor vehicle theft.

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

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

who are always assigned to a geographic district[25] and officer-supervisor relationships that lasted 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 rates for each district, separately for total crimes, property crimes, and violent crimes.[26]

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 nine years.

Relative to Police Officers, the racial makeup of supervisors[27] 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 eight years longer

[25] We 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.

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

[27] Recall *supervisors* are Sergeants who are officially assigned to conduct a Police Officer's annual evaluation.

than the average Police Officer. The average supervisor has 7.3 officers to evaluate every year, and the median number is seven. The 25th percentile is three officers, and the 90th percentile is 14 officers.

Table 1: Summary Statistics

	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.

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 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. White and Hispanic officers 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 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 third row lists the number of TRR filings, which is a proxy measure for use of force. The average officer files about 0.05 reports a month, equating to about 1.2 filings every two years. Black officers, however, files about half as many reports as white and Hispanic officers.

The remaining rows depict the number of monthly arrests by arrest type. For example,

Table 2: Racial Differences in Work Measures

	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)
TRR filings	0.05	0.05	0.03	0.06	-0.02 (0.000)	0.00 (0.039)
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)
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.

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 ($p < 0.01$). 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 -0.7 arrests per month or 8.4 fewer arrests per year ($p < 0.01$).

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 and 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 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 proactive arrests like drug and traffic arrests, which are proactive in that they are more likely to have originated from an officer-initiated incident. This classification of "proactive arrests", which allow for greater officer discretion, can also be seen as a delineation between appropriate and inappropriate uses of police authority.[28] 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. Relatedly, 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 Identifying Assumptions

This section outlines the empirical strategy to examine whether supervisors are less likely to acquire information about minority officers. We exploit two institutional features of the Chicago Police Department that allows us to estimate plausibly causal estimates of

[28] We 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.

the black-white recognition gap.

First, 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.[29] Although the vast majority of supervisor relationships last one year[30], 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 empirically test whether officers are as-good-as-randomly assigned to supervisors in the data.

Second, we exploit the randomized timing of an officer's annual evaluation. 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. Because start dates are determined by a lottery number, this means that the evaluation quarter is essentially randomly assigned across officers.[31] 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 (Q2).

The randomized timing of the annual evaluation in combination with the new supervisor assignment every January allows us to estimate plausibly causal estimates of the black-white nomination gap and how they evolve as supervisors learn more information about their officers due to the annual evaluation.

[29] About 96 percent of officers are assigned to a supervisor in January of each calendar year.

[30] Of all supervisor assignments between 2009 and 2015, 89 percent last exactly one year.

[31] After 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.

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.[32]

One way to examine the validity of this assumption is to test whether officers of different races are differentially likely to be assigned to a supervisor of a given race. If there is no supervisor-sorting, then we would expect white officers assigned to white supervisors to look similar to white officers assigned to black supervisors, and similarly for black officers.

Table 3 reports average lagged annual work 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, to black supervisors in column 2, and the p-value from a t-test of a difference in means after controlling for officer birth year, experience, unit, year fixed effects, and unit-year fixed effects.[33] Panel B lists the same for black officers, and Panel C lists the same for Hispanic 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.

[32] Further, we are able to control for officer work measures.
[33] Standard errors are clustered at the officer level.

Table 3: Officer Work Measures and Supervisor Race, Comparison of Means

	White Supervisor (1)	Black Supervisor (2)	p-value (3)
Panel A: White Officers			
Total Arrests	29.15	30.36	0.048
Violent-Crime Arrests	4.60	4.12	0.101
Property-Crime Arrests	3.96	3.23	0.000
Non-Index Crime Arrests	20.58	23.01	0.130
TRR Filings	0.61	0.59	0.056
Strong Force Ratio	0.88	0.91	0.738
Complaints	0.54	0.59	0.051
Panel B: Black Officers			
Total Arrests	16.57	14.47	0.027
Violent-Crime Arrests	3.80	3.25	0.000
Property-Crime Arrests	2.67	2.08	0.005
Non-Index Crime Arrests	10.11	9.14	0.173
TRR Filings	0.38	0.30	0.013
Strong Force Ratio	0.88	0.89	0.135
Complaints	0.53	0.51	0.025
Panel C: Hispanic Officers			
Total Arrests	28.94	27.91	0.016
Violent-Crime Arrests	5.20	4.31	0.000
Property-Crime Arrests	4.06	3.02	0.000
Non-Index Crime Arrests	19.67	20.58	0.093
TRR Filings	0.65	0.48	0.000
Strong Force Ratio	0.91	0.97	0.147
Complaints	0.58	0.54	0.013

Source: CPD analysis sample.
Notes: This table reports mean lagged annual work measures for officers assigned to white supervisors and black supervisors. The third column reports the p-value from a t-test of a difference in means after controlling for officer birth year and tenure, and including fixed effects for unit, year, and unit-year. Non-index arrests include arrests for non-property and non-violent crimes.

The table says that the average white officer assigned to a white supervisor this year had 29.2 arrests last year, of which 4.6 were for violent crimes, 3.96 were for property crimes, and 20.6 were for non-index crimes. The average white officer assigned to a black supervisor this year had 30.4 arrests last year, of which 4.1 were for violent crimes, 3.2 were for property crimes, and 23 were for non-index crimes. The average black officer assigned to a white supervisor this year had 16.6 arrests last year, of which 3.8 were for

violent crimes, 2.7 were for property crimes, and 10.1 were for non-index crimes. The average black officer assigned to a black supervisor this year had 14.5 arrests last year, of which 3.3 were for violent crimes, 2.1 were for property crimes, and 9.1 were for non-index crimes.

Overall, the numbers suggest that officers assigned to white supervisors vs. black supervisors look very similar in terms of work performance. Only one work measure for white officers has a statistically significant difference at the 1 percent level (two measures for black officers), and the difference is fewer than one arrest a year. Most importantly, the summary statistics suggest that if there is any selection, it would work against our findings. That is, black officers appear to be positively selected to white supervisors and white officers appear to be slightly negatively selected to white supervisors based on arrest numbers. This suggests that, in the absence of racial bias, white supervisors should nominate black officers more often than white officers based on their work performance measures. We also attempt to test for personality traits, such as penchant for aggressive behavior, by looking at the officer's rate of strong-force use. But even here, we do not see any statistically significant differences between officers assigned to white supervisors vs. black supervisors. Of course, we cannot test whether all 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, white 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 in Section 5.1 by leveraging the randomized quarter of evaluation.

5 Results

5.1 Black-White Gap by Evaluation Quarter

This section explores whether the black-white nomination gap is due to lack of information acquisition by supervisors. 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.

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 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. We estimate the following model, separately for white supervisors and black supervisors:

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

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 $RQ = 0$. The reference quarter is $RQ = -2$, or two quarters prior to the evaluation quarter.[34] 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, $RQ = 0$.

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 $RQ = -2$. If the black-white difference does not change in subsequent quarters, 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

[34] We drop observations that are three quarters before the evaluation quarter and three quarters after the evaluation quarter. This is because these observations are all either in January, February, and March (if three quarters prior) and October, November, and December (if three quarters post). Therefore, it is difficult to disentangle the calendar-month effect from the distance-to-evaluation effect.

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 cluster standard errors at the officer level.[35]

Table 4 reports estimates for δ^q (white officers), β_1^q (black-white difference), and β_2^q (Hispanic-white difference) for white supervisors in Panel A and for black supervisors in Panel B. White supervisors are more likely to nominate white officers as they move closer to the evaluation quarter, but this behavior stops afterward. Relative to two quarters before their evaluation, white officers are 0.68 percentage points (23.4 percent) more likely to be nominated in the quarter before their evaluation and 0.94 percentage points (32.4 percent) more likely to be nominated in their evaluation quarter. These estimates are statistically significant at the 1 percent level. After the evaluation, the relative nomination likelihood is 0.22 percentage points higher but not statistically significant. Two quarters post-evaluation, there is essentially no difference in nomination likelihood relative to two quarters prior.

For black officers assigned to white supervisors, the story is a different one. Although black officers are also more likely to be nominated in the quarter leading up to and including their evaluation, relative to two quarters prior, the relative increase is less than half the increase for white officers (specifically, about 35.4 to 42.7 percent less) and becomes negative after the evaluation. In fact, the estimates are stable around -0.4 to -0.5 percentage points across all quarters. Taken together, the relative increase in nomination likelihood suggests that white supervisors may be learning about their black officers, but the persistent, negative black-white disparity suggests that the learning is not manifesting in changed nominating behavior. This behavior is consistent with taste-based discrimination. By contrast, the nomination patterns for Hispanic officers are similar to those for white officers. In Appendix Table A4, we estimate a version with officer fixed effects

[35] As a robustness check, we estimate standard errors clustered at the supervisor level and robust standard errors. The results are essentially the same.

instead of supervisor fixed effects. The results are similar.

Table 4: Racial Difference in Nomination Likelihood by Quarter

Estimates for:	Outcome Variable: Nominated White Officer (1)	 Black-White Gap (2)	 Hispanic-White Gap (3)
Quarter relative to two quarters before evaluation			
Panel A: White Supervisors			
One quarter pre-evaluation	0.00683***	-0.00421	0.00133
	(0.00210)	(0.00282)	(0.00363)
Evaluation quarter	0.00937***	-0.00507*	0.00182
	(0.00213)	(0.00275)	(0.00351)
One quarter post-evaluation	0.00219	-0.00531*	-0.00203
	(0.00227)	(0.00270)	(0.00348)
Two quarters post-evaluation	-0.000280	-0.00425	-0.000953
	(0.00257)	(0.00304)	(0.00386)
Observations		154,964	
Panel B: Black Supervisors			
One quarter pre-evaluation	-0.000218	0.00450	0.0180**
	(0.00648)	(0.00697)	(0.00939)
Evaluation quarter	0.00127	0.00276	0.0227***
	(0.00631)	(0.00667)	(0.00905)
One quarter post-evaluation	-0.0129*	0.0139**	0.0245***
	(0.00588)	(0.00620)	(0.00847)
Two quarters post-evaluation	-0.00625	0.00450	0.0210**
	(0.00694)	(0.00701)	(0.0107)
Observations		26,556	
Baseline B-W Nomination Gap		-0.0045	

Source: CPD analysis sample.
Notes: This table depicts how the probability of nomination changes by quarter relative to two quarters before the officer's evaluation. Estimates are reported for white supervisors in Panel A and for black supervisors in 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. Standard errors clustered by officer are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We can examine black supervisors' behavior in Panel B as a contrast to provide context for white supervisors' behavior. The analysis suggests that black supervisors largely do not change their nomination behavior for white and black officers before and in the evaluation quarter. But, white officers are less likely to be nominated after their evaluation relative to

two quarters prior. The black-white gap widens by -1.3 percentage points (137 percent) in the quarter immediately following their evaluation, and this is significant at the 5 percent level.

To summarize, white supervisors nominate white and Hispanic officers at similar rates and black officers at a lower rate. These patterns suggest the black-white gap is not due to in-group favoritism towards white officers but rather bias against black officers.

5.2 Black-White Gap by Arrest Record

Next, we further examine whether the black-white nomination gap is due to bias against black officers or in-group favoritism towards white officers. Specifically, we ask whether an officer's arrest record affects a supervisor's likelihood of nomination and whether there are any differential effects for minority officers. That is, conditional on the officer's arrest record, is there a black-white disparity 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:

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

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. Standard errors are clustered by officer.[36]

The parameters of interest are β_1^c, which tell us how the nomination likelihood changes as the number of arrests last month increases, and β_2^c, which tell us how the black-white difference changes by the number of arrests. We expect β_1^c to be positive and increasing in the number of arrests. If the black-white gap in award nominations does not vary by the number of arrests, then β_2^c will be zero. As the baseline black-white gap is negative, a negative β_2^c indicates that the black-white gap widens with the number of arrests, whereas a positive β_2^c indicates that the black-white gap narrows with the number of arrests.

Table 5 reports estimates for white officers in column 1, the black-white difference in column 2, and the Hispanic-white difference in column 3. Panel A reports estimates for white supervisors, and Panel B for black supervisors. 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.

[36] When we estimate standard errors clustered at the supervisor level or estimate robust standard errors, the results are essentially the same.

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			
One arrest	0.00912***	-0.00497**	-0.000903
	(0.00135)	(0.00193)	(0.00239)
Two arrests	0.0147***	-0.00761***	0.000299
	(0.00190)	(0.00278)	(0.00329)
Three arrests	0.0216***	-0.0105**	0.000148
	(0.00260)	(0.00408)	(0.00447)
Four arrests	0.0250***	-0.00252	-0.000720
	(0.00326)	(0.00586)	(0.00566)
Five or more arrests	0.0566***	-0.0236***	-0.0156**
	(0.00256)	(0.00481)	(0.00428)
Observations		171,094	
Mean Pr(Nom) for White Officers		0.031	
Panel B: Black Supervisors			
One arrest	-0.000465	0.00354	0.00734
	(0.00309)	(0.00354)	(0.00639)
Two arrests	0.00855	-0.00635	0.000517
	(0.00509)	(0.00570)	(0.00997)
Three arrests	0.0101	-0.000252	0.00940
	(0.00659)	(0.00800)	(0.0137)
Four arrests	0.00844	-0.0151	0.00810
	(0.00889)	(0.0102)	(0.0171)
Five or more arrests	0.0451***	-0.00294	-0.0224*
	(0.00763)	(0.0113)	(0.0125)
Observations		29,413	
Mean Pr(Nom) for White Officers		0.022	

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 complaints, lagged overall crime rate, lagged violent crime rate, and the share of black supervisees. Standard errors clustered by officer are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The black-white nomination gap widens as the number of arrests increases. Among officers with five or more arrests, the relative black-white difference widens by 2.4 percentage points ($p < 0.01$). 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. In Appendix Table A5, we estimate a version with officer fixed effects instead of supervisor fixed effects. The results are very similar.

We also examine whether white supervisors are less likely to nominate Hispanic officers, another racial minority in the Chicago Police Department. The Hispanic-white difference is pretty trivial and not statistically significant until the five or more arrests category (Panel A, column 3). Among officers with at least five arrests, the Hispanic-white gap in nomination probability widens by 1.6 percentage points ($p < 0.05$).

When comparing between the two racial minorities, the black-white difference is statistically significantly different from the Hispanic-white difference among officers with one to three arrests and not for those with four or more arrests. This suggests that white supervisors are less likely to nominate black officers relative to 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 higher than average arrests.

We examine black supervisors' behavior in Panel B. Most of the point estimates are not 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. One exception is the Hispanic-white gap of -0.022 among officers with five or more arrests ($p < 0.1$).

To summarize, we find that black officers are nominated less frequently than white officers even conditional on arrests. The gap widens as the number of arrests increases, which is the opposite of what one would expect if the disparity were due to white supervisors' inaccurate beliefs about black officers' work measures. This is supported by the fact that we do not see similar behavior exhibited by black supervisors. Instead, the results suggest that the source of discrimination may be taste-based. These findings are consistent with

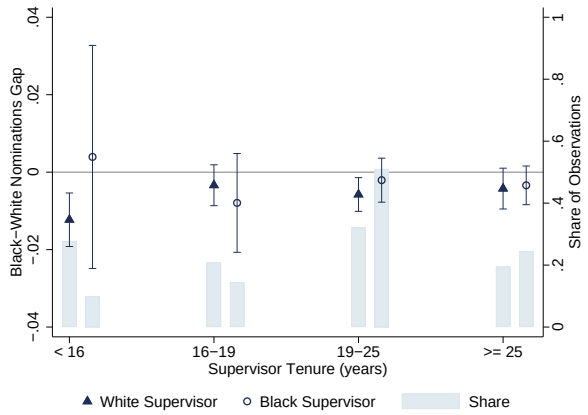
the analysis from the previous section, which finds that supervisors may not be updating their behavior as they learn more information.

5.3 Black-White Gap by Supervisor Characteristics

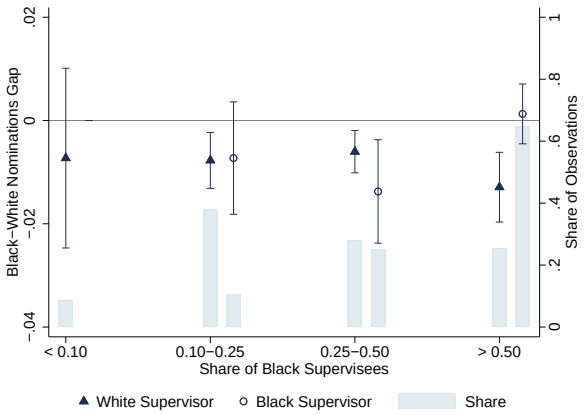
We also examine how the black-white nomination gap changes by supervisor characteristics in Figure 1. Panel A reports the black-white nomination gap by supervisor tenure, with quantiles defined by the 25th, 50th, and 75th percentile values (about 16 years, 19 years, and 25 years, respectively). The patterns indicate that less experienced white supervisors have the largest negative black-white nomination gap at -1.2 percentage points ($p < 0.01$). In contrast, black supervisors with similar levels of experience have a positive black-white gap but this is not statistically significant.

Figure 1: Black-White Nomination Gap by Supervisor Characteristics

(a) Supervisor Tenure



(b) Share of Black Supervisees



Notes: This figure depicts how the black-white nomination gap changes by supervisor tenure (Panel A) and by the share of black supervisees (Panel B). 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, the number of supervisees, and the share of black supervisees. Wings depict 95% confidence intervals with standard errors clustered at the officer level.

Panel B reports the black-white nomination gap by the share of black supervisees. Here, the patterns indicate that the black-white gap among white supervisors becomes more negative and statistically significant when the share of black supervisees exceeds 50 percent.

Among supervisors with a majority of black officers, white supervisors have a black-white nomination gap of -1.3 percentage points ($p < 0.01$), while it is slightly positive and not significant for black supervisors.

6 Experimental Evidence

The previous section presented evidence that white CPD supervisors are less likely to acquire information about their black officers, and this may be due to taste-based discrimination. To more concretely test this, we ran an online experiment to measure the review process in the nomination decision.[37] This section discusses the evidence on whether evaluators choose to engage with minority officers.

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 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.[38]

[37] The experiment was pre-registered in the AEA RCT Registry, AEARCTR-0005929.

[38] In 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.

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.[39] 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 (race, sex, and age) and participants had to mouse over a profile to reveal full information about the officer.[40] 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, three of the four profiles were always white officers and the race of the fourth profile was randomly chosen amongst white, black, and Hispanic. In the second 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. The ordering of these two iterations was randomized.

The ordering of the two tasks was randomized, and the display ordering of officer profiles in each of the tasks was also 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 compari-

[39] See Appendix Figure B2 for a screenshot of the task.
[40] See Appendix Figure B3 for a screenshot of the task.

son).[41] 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; participants 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.[42] We recruited 411 MTurk workers (hereafter "evaluators") 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 evaluators paid attention during the survey. For the analysis, we decided to include evaluators who passed at least two of the three data quality checks. This restriction reduces our final analysis sample to 407 evaluators.

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

First, we seek to replicate the results from the CPD administrative data and ask whether black officers were less likely to be nominated for an award. Columns 1 through 3 of Table 6 report results from the pairwise comparison of a black and non-black (white or Hispanic)

[41] These 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.

[42] It 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.

officer. Column 1 reports results from all MTurk evaluators, column 2 is restricted to white MTurk evaluators, and column 3 is restricted to black MTurk evaluators.

Table 6: Impact of being Black on Nomination Likelihood

Outcome Variable: Nominated				
Pairwise Comparison:	Black v. Non-Black			High v. Low
Race of MTurk Worker:	All	White	Black	All
	(1)	(2)	(3)	(4)
High-Quality Profile	0.483***	0.410***	0.771***	0.524***
	(0.0921)	(0.0947)	(0.284)	(0.169)
Black Officer	-0.0883***	-0.103***	-0.0163	-0.0324
	(0.0311)	(0.0361)	(0.0823)	(0.119)
High-Quality x Black Officer	0.0326	0.0487	-0.00390	0.141
	(0.0448)	(0.0518)	(0.121)	(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 the screen. Officer traits include officer age, tenure, arrests, and complaints. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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 evaluators, we see that the results are driven by white evaluators (column 2). The black-white gap among white evaluators 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 evaluators. These results are consistent with the CPD analysis, which found the

black-white nomination gap widens for white supervisors.

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 evaluators were able to discern the quality difference between the two officer profiles. Interestingly, when the officers are both black or both female, evaluators 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.

6.4 Do evaluators choose to engage with black officers?

Table 7 presents summary statistics on MTurk evaluators' engagement with white officers, black officers, and Hispanic officers. We examine three different measures of engagement: ever moused over (Panel A), first mouseover (Panel B), and mouseover duration measured in seconds (Panel C). Table 8 reports the findings in a regression framework, where we control for the profile location on the computer screen, and the evaluator's starting mouse position.

The first row in Panel A tells us that evaluators 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 black-white difference is borderline significant, with a p-value of 0.107. The Hispanic-white difference is not statistically significant.

When the officer pool is predominantly white—an environment that resembles the

Chicago Police Department—the black-white engagement gap widens. Black officers are 7.3 percentage points less likely to be moused over compared to white officers ($p < 0.05$). However, if the minority officer is Hispanic, then there is no statistically significant 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 racial difference in the amount of time spent reviewing profiles (Panel C). Evaluators spend around half a second more reviewing black and Hispanic profiles, and these are significant at the 1 percent level.

Table 7: Evaluator Engagement 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)
Predom. White Officer Group	83.8%	76.6%	86.7%	-0.073 (0.036)	0.028 (0.420)
Het. Race Officer Group	85.2%	82.7%	80.6%	-0.025 (0.276)	-0.046 (0.054)
Panel B: Outcome Variable: First Mouseover					
All Officers	30.0%	31.9%	28.4%	0.019 (0.419)	-0.016 (0.490)
Predom. White Officer Group	28.9%	35.7%	35.6%	0.069 (0.155)	0.067 (0.152)
Het. Race Officer Group	32.8%	31.0%	26.5%	-0.018 (0.581)	-0.063 (0.048)
Panel C: Outcome Variable: Mouseover Duration (seconds)					
All Officers	2.33	2.78	2.89	0.448 (0.000)	0.559 (0.000)
Predom. White Officer Group	2.36	3.40	3.07	1.041 (0.000)	0.714 (0.001)
Het. Race Officer Group	2.27	2.64	2.85	0.366 (0.006)	0.576 (0.000)

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.

When the officer pool is heterogeneous—that is, white, black, and Hispanic officers are represented in equal numbers—the black-white disparity disappears and an Hispanic-white disparity emerges. Hispanic officers are 4.6 percentage points less likely to be moused over and 6.3 percentage points less likely to be the first mouseover, relative to

white officers. One potential explanation is that when the two racial minorities (black officers and Hispanic officers) are in the same comparison pool with white officers, black officers crowd out Hispanic officers in regards to evaluator attention. 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 evaluators' decisions on who to mouseover first. This, however, does appear to be a crowd-out effect because when the black officers are the sole minority officer in a group with three white officers, evaluators are less likely to engage with them.

Table 8: Impact of Officer Race on Evaluator Engagement

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

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 evaluator's starting mouse position. Standard errors are in parentheses. Panels A and B report robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6.5 The importance of evaluator engagement

In this section, we examine how evaluator engagement affects the probability of nomination. Table 9 reports the nomination likelihood conditional on officer race and the engagement measure: ever moused over (Panel A), first mouseover (Panel B), and mouseover duration (Panel C). All estimates control for officer traits, the officer's profile location on the screen, and the evaluator's starting mouse position on the screen.

Profiles that were moused over are on average 8.8 percentage points more likely to be nominated (Panel A). This suggests that mouseover activity is a good measure of the evaluator's interest in the officer. Workers who engage with 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. Column 2 reveals that these results are driven by the predominantly white officer pool.

We also analyze first mouseover as a proxy for the intensity of an evaluator's interest in an officer. 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 profiles, however, black profiles who are moused over first are 9.2 percentage points even more likely to be nominated ($p < 0.1$). Column 2 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 an evaluator chooses to learn about the minority officer first, then the evaluator 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 imply 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.

It is also informative to examine the situation when evaluators do not engage with officers. In this instance, we may expect no racial difference in nomination likelihood since evaluators have no work information about the specific officers. However, we find that black officers are less likely to be nominated relative to white officers whose work measures were also not revealed. This difference widens to -38 percentage points and becomes statistically significant at the 1 percent level when the officer pool is predominantly white.

It is difficult to ascertain whether this behavior is driven by distaste or inaccurate beliefs about black officers (for example, evaluators may believe black officers are less productive

than white officers and therefore may not engage with them in the interest of time). The estimates in Panel B may shed some light. Among officers who were moused over, but were not the first mouseover, black officers were still less likely to be nominated relative to their white counterparts. In a predominantly white officer pool, black officers who were not the first mouseover are 13.8 percentage points less likely to be nominated than their white peers who were not the first mouseover. In contrast, Hispanic officers who were not the first mouseover are 19.4 percentage points more likely to be nominated relative to their white peers. Both of these differences are significant at the 10 percent level. These results suggest that evaluators are not updating their information about black officers, which is consistent with our results using CPD administrative data in Section 5.1.

Next, we consider the length of the review process (Panel C). The longer an evaluator spends viewing an officer's profile, the higher the chance of a nomination: an increase of 4.4 percentage points for each additional second ($p < 0.01$). 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. When the officer pool is predominantly white, 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 baseline black-white nomination gap of -0.198, this suggests that an evaluator would need to spend about five additional seconds–more than two standard deviations–on a black officer's profile in order to equalize the nomination probability for white vs. black officers. In contrast, any Hispanic-white differences become insignificant in a predominantly white officer pool.

Taken together, the results appear to suggest that there is a dichotomy in the type of evaluators who are likely to engage with minority officers vs. not. Because all officer profiles are of average quality, the decision to nominate an officer should be independent of

officer performance and instead reflect solely the evaluator's preferences. Some evaluators are interested in engaging with black and Hispanic officers, and those evaluators are also more likely to nominate minority officers. This suggests that black officers may benefit from having supervisors who are interested in interacting with them.

On the other hand, some evaluators are less interested in engaging with black officers. Black profiles are less likely to be moused over than white profiles, on average, and this is more salient when evaluators are choosing among three white officers and one black officer. We do not see similar patterns when evaluators are choosing among three white officers and one Hispanic officer. Further, black officers are less likely to be nominated than white officers conditional on neither being the first profile the evaluator moused over.

It is informative to compare estimates when the officer pool is predominantly white vis-a-vis evenly distributed among white, black, and Hispanic officers. Most estimates lose statistical significance, but the point estimates remain similar for the most part; black officers continue to be penalized more than white officers from a lack of evaluator engagement. What disappears when we move to a more diverse officer pool are the *benefits* black officers had received from greater evaluator engagement. For example, the black-white difference among profiles that were the first mouseover becomes negative (from 0.19 to -0.01). This suggests that having a diverse police force may not actually eliminate the black-white recognition gap. This implication is also consistent with our finding using CPD administrative data, that the black-white nomination gap becomes more negative among white supervisors with a majority of black supervisees. These findings underscore the importance of addressing the black-white *promotion* gap, as simply diversifying incoming police officers may be limited in its efficacy.

Table 9: Impact of Engagement on Nomination Likelihood

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

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 evaluator's starting mouse position. Officer traits include officer age, tenure, arrests, and complaints. Standard errors are in parentheses. Panel A reports robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

7 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)[43], 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.[44]

A potential solution that has been put forth is to increase racial and gender diversity among officers, who are traditionally homogeneous.[45] 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).[46] 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

[43]Mueller-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.

[44]The release of detailed administrative data has allowed for research on various important questions, such as crime reduction (Blanes i Vidal and Kirchmaier, 2018; Mastrobuoni, 2020).

[45]For 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/)

[46]McCrary (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.

Police Department in 2015, and 56 percent of those at the rank of Sergeant or higher.[47] In addition to improving policing quality[48], 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 hinder minorities' career progression.[49] 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 randomized timing of the evaluation, and find that white supervisors are more likely to nominate officers in the evaluation quarter, suggesting that information acquisition is important for career recognition, but also that black officers benefit less than their white peers, suggesting that white supervisors are not updating.

To supplement our CPD analysis, we run an online experiment on Amazon Mechanical Turk and, again, find that black officers are less likely to be nominated than their non-black peers. In terms of engagement, we find that black profiles are the least likely to be moused over, and that the black-white gap more than doubles when evaluators are choosing among three white officers and one black officer.

[47] These numbers do not include civilian Department members.

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

[49] For example, Egan et al. (2018) find that female financial advisors are 20 percent 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.

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 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.

Blanes i Vidal, J. and T. Kirchmaier (2018). The effect of police response time on crime clearance rates. *The Review of Economic Studies 85*(2), 855–891.

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.

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.

Mastrobuoni, G. (2020). Crime is terribly revealing: Information technology and police productivity. *The Review of Economic Studies 87*(6), 2727–2753.

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.

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.

Weisburst, 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.