

The Black-White Recognition Gap in Award Nominations*

Nayoung Rim[†]

Roman Rivera[‡]

Andrea Kiss[§]

Bocar Ba[¶]

September 8, 2021

There is substantial evidence showing racial bias in firms' hiring decisions, but less is known about bias in career progression. We construct a dataset of police award nominations to estimate the Black-white nomination gap. Exploiting institutional features of the police department to obtain plausibly causal estimates, we find that white supervisors are less likely to nominate Black officers relative to other officers conditional on their work performance. We also find suggestive evidence that supervisors are less likely to learn about and advocate for Black officers. These results, corroborated by an online experiment, suggest that the Black-white nomination gap may stem in part from a racial engagement and advocacy gap. Given the reliance on subjective evaluations for promotions in many organizations, our findings may have important implications for the Black-white promotion gap and the lack of diversity in upper-management positions.

JEL Codes: J71, M51, J45, J48

Keywords: racial bias, discrimination, promotions, supervisors, police, evaluations

*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., the Huettel Laboratory, and many conference and seminar discussants 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.

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

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

[§]Carnegie Mellon University, Department of Social and Decision Sciences. Email: akiss@andrew.cmu.edu

[¶]Duke University, Department of Economics. Email: bocar.ba@duke.edu

1 Introduction

Most organizations have promotion processes that are based on subjective evaluations of an employee's work performance. This may foster an environment where, due to the importance of having advocates and connections, racial or gender disparities can arise in career outcomes. This advocacy gap may explain why male workers are promoted faster if they are assigned to male managers vs. female managers (Cullen and Perez-Truglia, 2020) and why Black employees are less likely than their white colleagues to report receiving support for advancement at work (McKinsey & Company, 2021).

While the economics literature on racial discrimination in the labor market has largely focused on firms' hiring and interviewing decisions,¹ less is known about racial bias in career recognition and progression. Given that a majority of managers are white and connections can engender career opportunities,² does the Black-white promotion gap exist partly because managers are not advocating equally for their Black and white colleagues?

This question is challenging to answer empirically because (1) it requires data on workers' performance, which are often hard to measure and difficult to collect due to its proprietary nature, (2) it requires data on subjective evaluations, which are also typically proprietary, and (3) supervisor-supervisee relationships are often endogenously formed, making it challenging to identify a causal estimate of the racial gap in promotions.³

We overcome these challenges by constructing a panel dataset of Chicago Police Department (CPD) officers between 2009 and 2015 that contains detailed personnel information on arrests, use of force, misconduct, and supervisor assignments. Using supervisor nominations for departmental awards as our outcome measure, we examine whether there is a racial gap in career recognition in the second largest police department in the US. CPD

¹See, for example, Bertrand and Mullainathan (2004); Kessler et al. (2019); Neumark et al. (2019).

²According to Jobvite.com, 46 percent of hires were from referrals despite referrals making up only seven percent of all applicants. See <https://theundercoverrecruiter.com/infographic-employee-referrals-hire/>

³For example, we may be concerned that higher performing white workers and/or lower performing Black workers may sort to white managers. In this case, we would see a racial gap in promotions even in the absence of racial bias.

awards are a measure of career recognition and used in important decisions related to career advancement, such as merit promotions.⁴ Importantly, award nominations are *subjective* evaluations of an officer’s work performance in that a supervisor singularly decides whether or not to nominate an officer based on the officer’s performance.⁵

We ask (1) whether white supervisors are less likely to nominate Black officers relative to other officers conditional on their work performance, and (2) whether white supervisors are biased in terms of whose performance they choose to learn about (a racial learning gap). In our setting, supervisors do not necessarily observe the officer’s day-to-day activities, making it costly for them to learn about the officer’s work performance. As a result, work-related evaluations like the decision to nominate an officer for an award, require effort. If the cost of acquiring information differs by race, then a racial gap in career recognition and progression may arise. Although our application focuses on law enforcement, this organizational structure (autonomous workers operating within a hierarchical organization) is common across all industries.

Our identification strategy exploits several institutional features of the CPD that allow us to obtain plausibly causal estimates of the Black-white recognition gap. First, officers are assigned to a new supervisor every January. Importantly, supervisor assignment is not a function of officer performance and is plausibly as-good-as-random conditional on officer district and year. This allows us to answer the causal question of whether white supervisors nominate Black and white officers at equal likelihoods.

⁴Merit promotions, which rely on a variety of officer qualifications, allow officers to move up the promotion list even if their examination score did not place them at the top of the list.

⁵We choose to focus on nominations as opposed to award receipt because the final decision is made by a higher-ranking officer or external Awards Committee, which introduces a layer of complexity and may introduce an additional source of bias into the decision process. At the same time, however, our decision to focus on nominations allows us to speak somewhat to actual award receipt as 71 percent of nominations are approved. 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 tenure. 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. Despite the inability to consider promotion and wage disparities, we believe a major contribution of this paper is its access to data on award nominations and the ability to examine how subjective evaluations may differ by race—a process that is largely a black box due to the lack of data.

We find that white supervisors are less likely to nominate Black officers relative to white officers, even conditional on the number of arrests. Award nominations are typically related to a reported crime, which can explain why an officer’s arrest record has a large impact on the likelihood of being nominated for an award.⁶ But, conditional on the number of arrests, Black officers are 28 to 32 percent less likely than white officers to be nominated. The Black-white nomination gap widens as the number of arrests increases, which is the opposite of what one would expect if the disparity were due to a supervisor’s (negative) beliefs about Black officers’ work performance (i.e., statistical discrimination).⁷

Our second research question asks whether the nomination gap exists because white supervisors are choosing to learn about white officers over black officers. To examine whether there is a learning gap, we leverage another institutional feature: supervisors are required to conduct annual evaluations of their assigned officers. Although there may be little interaction between officers and supervisors on a daily basis—and, thus, why information acquisition is costly—we assume that supervisors are motivated to gather information and learn about an officer’s work record for the annual evaluation. Further, officer evaluations are randomly assigned across the calendar quarter, which allows us to disentangle any seasonal trends in nomination decisions from the effect of information gathering.⁸ This analysis can also provide some insight into the source of the racial disparity. If the Black-white nomination gap is a result of statistical discrimination, then we would expect the gap to shrink around the time of the evaluation as supervisors learn more about their officers’ work records. In contrast, if the gap were due to racial bias, then we would expect the gap to persist.

⁶For example, white officers with more than three arrests per month (roughly the top 10 percent) experience a three-fold increase in nomination likelihood relative to officers with zero or one arrests.

⁷For example, statistical discrimination may arise if productivity is unobserved and one group is perceived to have a lower average productivity either through exogenous factors (Phelps, 1972) or through a self-fulfilling prophecy (Arrow, 1973), or if the two groups’ productivity distributions have different variances (Aigner and Cain, 1977).

⁸The evaluation quarter is essentially randomly assigned across officers because the evaluation must take place during the quarter prior to the quarter in which the officer joined the Department, and start dates are determined by a lottery number.

We find that white supervisors are more likely to nominate all of their officers in the evaluation quarter relative to the baseline quarter, suggesting that preparing for the annual evaluation does induce white CPD supervisors to learn about their officers. However, Black officers are consistently less likely to be nominated than white officers (15.5 to 42 percent less). This suggests that learning additional information does not translate into more nominations for Black officers as it does for white officers—even conditional on their arrest records. Furthermore, nominations for Black officers immediately revert to their baseline nomination likelihood after the evaluation quarter. In contrast, white officers continue to receive more nominations *after* the evaluation quarter, suggesting there may be longer lasting bonds between white supervisors and white officers.

Our preferred interpretation of the results is that though a learning gap may exist, it cannot fully explain the nomination gap. Instead, we posit that the disparity can be explained by an advocacy gap: supervisors are less likely to choose to advocate for (nominate) Black officers relative to white officers, even after learning about their work performance. For example, although CPD supervisors are more likely to nominate black officers in the evaluation quarter relative to the baseline quarter, the Black-white nomination gap is the largest in the evaluation quarter and the quarter immediately following it.

This interpretation is supported by the results of an online experiment, wherein we experimentally examine whether evaluators choose to learn about minority officers. We ask Amazon Mechanical Turk (MTurk) workers to evaluate officer profiles and to nominate one for an award. In one 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 the least likely to be moused over, indicating a racial engagement gap. In another task, evaluators choose between a Black officer and a non-Black officer. Although officer performance levels are randomly determined and visible (e.g., not covered), evaluators are 5.6 to 8.8 percentage points less likely to nominate Black officers over white officers. Since a racial

disparity exists even in the presence of full information, these results are not consistent with statistical discrimination.

Although we cannot definitively determine the source of the discrimination,⁹ our findings suggest that we can rule out statistical discrimination as the source. Because CPD supervisors and MTurk evaluators are basing their nomination decisions on actual performance measures as opposed to a noisy signal, our finding of a negative Black-white nomination gap cannot be explained by a lack of accurate information. As further support, the Black-white gap *widens* as the number of arrests increases and in the evaluation quarter. Therefore, the fact that white CPD supervisors and MTurk evaluators are less likely to engage with, learn about, and advocate for Black officers relative to white officers, does not appear to be from a lack of relevant information.

Our paper relates to the literature on social networks in the workplace. Prior research has documented the importance of homophily in the workplace. Physicians prefer to refer patients to specialists of the same gender (Sarsons, 2019; Zeltzer, 2020). Market traders are more likely to follow the trades of own-race colleagues (Levine et al., 2014). We find that CPD supervisors and MTurk evaluators, both groups that are largely white, are more likely to nominate white officers over Black officers. This is consistent with studies that find the importance of own-race or own-sex matching with supervisors for career outcomes (Cullen and Perez-Truglia, 2020; Giuliano et al., 2009, 2011).

Our paper is also consistent with studies that find that minorities are less likely to be acknowledged for their work (Hengel, 2019; Lamont, 2018; Levine et al., 2021; Sarsons, 2021).¹⁰ We find that white supervisors are less likely to learn about Black officers, leading to a racial disparity in award nominations. This racial attention gap has been empirically documented in various settings, including the labor market and the rental market (Bartoš et al., 2016; Levine et al., 2021). More broadly, our paper contributes to the literature

⁹We are unable to do so because we do not have information on beliefs and preferences.

¹⁰Hengel (2019) and Sarsons (2019) find that female minorities are less likely to be acknowledged for their work, while the others focus on racial minorities.

that establishes the existence of bias among managers and work colleagues (Bertrand and Mullainathan, 2004; Egan et al., 2020; Glover et al., 2017; Sarsons, 2019).

With respect to law enforcement, our study adds to the growing research on racial bias in policing.¹¹ Rather than focusing on racial bias in officer-civilian interactions, we examine racial bias within police departments. Our findings suggest that racial issues in policing are not just at issue between the police and the public, but also within departments and, thus, that simply hiring minority officers may be limited in its efficacy.

We begin the rest of the paper with a short description of CPD’s organizational structure and the awards process (Section 2). Section 3 describes our data and presents summary statistics. Section 4 discusses the identifying assumptions. We present results using administrative CPD data in Section 5 and the experimental evidence in Section 6. Section 7 concludes.

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.¹² 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 Accountability Task Force, 2016).¹³

¹¹See, 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); Goncalves and Mello (2020); Hoekstra and Sloan (2020); Horrace and Rohlin (2016); Knowles et al. (2001); Mason (2007); Nix et al. (2017); Rim et al. (2020); Weisburst (2018); West (2018).

¹²Between 2012 and 2014, three districts were dissolved leaving 22 geographic districts.

¹³Additionally, officers may bid for district transfers when vacancies occur. Successful bidders are chosen based on their seniority.

The most common rank in the CPD is that of *Police Officer*. This is distinct from that of a detective or a specialized officer. Police officers generally work in one of the 25 geographic districts (each of which corresponds to a unit), where they patrol, respond to 911 calls, and perform proactive policing activities. They are supervised on a day-to-day basis by sergeants who are working the same shift (day, watch time, and unit). Sergeants' daily responsibilities include participating in roll call, supervising criminal investigations (e.g., protecting the scene, establishing the perimeter), and ensuring officers carry out their responsibilities.¹⁴

Sergeants are also tasked with conducting performance evaluations of a specific group of police officers, who are assigned annually. These officers typically serve in the same district and watch time as each other but, because of the rotational schedule, do not generally work together every day.¹⁵ To clarify, a *supervisor* by our definition differs from the *sergeant on duty*, who conducts roll call and ensures officers carry out their responsibilities. As an example, say Police Officer A is assigned to work a shift with Police Officer B, Police Officer C, and Sergeant S. Further, suppose Sergeant S was not assigned to conduct Officer A's annual evaluation in January. Sergeant S, who is Officer A's *sergeant on duty*, conducts roll call and supervises criminal investigations but in our analysis, we do not consider Sergeant S to be Officer A's *supervisor*.

All CPD officers work on a rotational schedule, where they rotate their off-days each week. A notable feature of this system is that duty cycles do not occur on a weekly basis such that officers are not consistently assigned to work with the same group of officers or sergeants.¹⁶ The rotational system of scheduling and other features of the CPD make it costly for supervisors to learn about their officers' work performance. Indeed, a 2017 DOJ report found this system "prevents supervisors from establishing mentoring relationships

¹⁴Section 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.

¹⁵According to shift-assignment data, 62 percent of officer shifts between 2010 and 2019 were not with the officer's annually assigned supervisor.

¹⁶A duty cycle is typically four on-duty days followed by two off-duty days.

with officers and providing guidance” (U.S. Department of Justice, 2017, p. 108). Even if officers were consistently assigned to work with the same sergeant, sergeants are required to “spend too much time doing non-supervisory tasks at the expense of providing officers supervision” (U.S. Department of Justice, 2017, p. 107).

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. However, conversations with CPD command staff revealed that CPD supervisors “do not understand how the PRS works or how to use the information it presents” (U.S. Department of Justice, 2017, p. 112).

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 evaluations¹⁷ and merit promotions¹⁸.

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. Nearly 90 percent of nominations for police officers are from sergeants. Of all nominations, 38.5 percent originate 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. An “incident” is not technically defined, but 85 percent of all nominations are related to a reported crime. There is no restriction on the number of times an officer may be nominated, as long as the nominations are for different

¹⁷Chicago Police Department, Career Development Directive, Employee Resource E05-01, Section IV.H., available at <http://directives.chicagopolice.org>.

¹⁸Section III.E.2, Employee Resource E05-05, available at <http://directives.chicagopolice.org>.

incidents.¹⁹ Supervisors are also not restricted in the number of award nominations they are allowed to submit. Our analysis focuses on nominations by annual supervisors as opposed to all sergeants to leverage the supervisor quasi-random assignment in order to recover a causal estimate of the Black-white nomination gap.

Nominations for department awards are forwarded to a higher-ranking officer or board for review and a final decision. Because the final decision depends on external reviewers, we focus our analysis on award *nominations*, which are singularly the nominator's decision. As we are interested in whether white supervisors advocate for Black officers and white officers equally, our outcome variable should capture the supervisor's desire or intent to nominate an officer regardless of the final award status. At the same time, however, our decision to focus on nominations allows us to speak somewhat to actual award receipt as 71 percent of nominations are approved.

As mentioned before, awards are important accomplishments that are related to career milestones, such as promotions. Most officers are promoted to sergeant via written examinations, which they are eligible for after five years of service. But officers can also be promoted to sergeant through a merit promotion, which are capped at 30 percent of all sergeant promotions. Merit promotions allow officers to move up the promotion list even if their examination score did not place them at the top of the list, through a nomination from a high-ranking member of the CPD command staff. Merit promotions rely on a variety of officer qualifications, such as awards.²⁰ For example, officers promoted to sergeant had nine more cumulative awards than the average police officer of equal tenure.

¹⁹There are a few exceptions to this rule. Recipients of the Honorable Mention Certificate may also be nominated for another department award for the same incident. Further, the 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>).

²⁰Merit promotions were created in response to complaints that the CPD promotional exams were discriminatory and in an effort to diversify senior officers. However, many officers currently feel that merit promotions are based on political connections rather than true merit—highlighting the importance of networks in career outcomes. See <https://www.npr.org/local/309/2019/12/11/787040792/beck-suspends-controversial-merit-promotions-in-police-department>.

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.²¹ Officer rank is taken from salary data provided by the Chicago Department of Human Resources (DHR), covering 2002 to 2017. 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.

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 nomina-

²¹Not all demographic information is complete in each file, so an aggregation of demographic variables across multiple files is necessary for identifying a unique officer. Over 99 percent of officers are matched to a unique gender, race, and appointment date.

tions 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).²²

Arrests The arrests dataset contains information on all arrests made by police officers. The dataset includes detailed information about the subject, crime, arrest location, and time of arrest. 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.²³ 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 they are more serious offenses.²⁴ Non-index crimes capture crimes that are less serious such as municipal code violations, traffic violations, and drug abuse violations.²⁵

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.

Use of Force 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. Using CPD’s guidance on the appropriate use of force in different situations, we identify officer actions that are categorized as “strong use of force.” Strong uses of force involve elevated levels of force that are generally intended to enact

²²An 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.

²³In 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.

²⁴Violent index crimes are murder, aggravated assault/battery, robbery, and rape. Property index crimes are larceny theft, burglary, arson, and motor vehicle theft.

²⁵A comprehensive list of crime categories can be found at http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html.

harm on or injure the subject.²⁶

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 have a district (unit) assignment. We focus on years 2009 to 2015 to maximize overlap across the different datasets. We further restrict our sample to officers at the rank of Police Officer who are always assigned to a geographic district²⁷ 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 are open to all Department members.²⁸ Appendix Table A1 provides a description of these awards. Almost all awards are Honorable Mention Certificate Awards, which are relatively broad in scope, compared to, say, the Police Blue Star Award for an officer injury, and often relate to a single incident (e.g., arrest) rather than an officer's body of work, which is the focus of other awards like the Chicago Police Leadership Award. This facilitates our empirical strategy of conditioning on arrests.

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 (per 1,000 residents) for each district, separately for total crimes, property crimes, and violent crimes.

²⁶Strong use of force may or may not necessarily involve weapons. Examples of strong use of force without weapons are take-downs, kicks, and punches.

²⁷We remove the three districts that closed between 2012 and 2014 (13, 21, and 23) from our analysis sample because we do not have crime statistics for these districts.

²⁸Most 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.

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 supervisors²⁹ 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 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

²⁹Recall *supervisors* are sergeants who are officially assigned to conduct a police officer's annual evaluation.

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)
Won	2.3%	2.8%	1.1%	2.9%	-1.6 (0.000)	0.2 (0.051)
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 in our analysis sample. The 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.

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 monthly probability of winning an award. It is very similar to the nomination numbers because most award nominations (about 71 percent) are approved.

The third 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 fourth 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, file about half as many reports as white and Hispanic officers.

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

the average officer makes 1.8 arrests every month. White and Hispanic officers are slightly over this average at 2 and 2.2 arrests, respectively, while Black officers are below this average at 1.2 arrests. The Black-white difference equates to 10 fewer arrests a year ($p < 0.01$). 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.³⁰ In Table 2, we see that white officers are about 2.4 to 2.6 times more likely than Black

³⁰We 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.

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 nominate minority officers conditional on their work performance and whether supervisors are biased in terms of whom to learn about when deciding whom to nominate. We exploit two institutional features of the Chicago Police Department that allow us to estimate plausibly causal estimates of 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.³¹ Although the vast majority of supervisor relationships last one year (78 percent), we may be concerned that some officer-supervisor relationships may have been arranged outside of the dispassionate 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 Section 4.1, 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 to examine how knowledge about an officer’s work performance may influence a supervisor’s nom-

³¹About 96 percent of officers are assigned to a supervisor in January of each calendar year.

ination likelihood. 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.³² Appendix Table A2 lists the evaluation quarters and evaluation due dates by start month. For example, if an officer started police academy in July (Q3), then his annual evaluation must take place in the second quarter of every calendar year (Q2).

To estimate the causal impact of information gathering, it is important for the evaluation period to differ across officers. This will allow us to control for any seasonal trends in nomination decisions. For example, nomination likelihoods may be highest in the summer, when relatively more crimes are being committed, and lowest in the winter, when relatively fewer crimes are being committed. If the annual evaluation differs across officers—and importantly, differs due to a random process—then we will be able to disentangle any seasonal trends from the true estimate of information gathering.

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

4.1 Exogeneity of Supervisor Assignment and Officer Performance

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

³²After passing a written exam, all CPD candidates are placed on an eligibility list according to a randomly assigned lottery number and called off in lottery order to enroll in the police academy.

be the result of officers sorting based on work performance measures. We argue that this sorting concern is mitigated in our setting due to the annual as-good-as-random re-assignment of supervisors.

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 white supervisor. If there is no sorting, then we would expect white officers and Black officers assigned to white supervisors to look similar to each other. We check this by estimating the following regression model:

$$\begin{aligned} WhiteSup_{it} = & \beta_0 + X'_{it}\beta + Z'_{i,t-1}\alpha_1 + (B_i \times Z'_{i,t-1})\alpha_2 + (H_i \times Z'_{i,t-1})\alpha_3 \\ & + (A_i \times Z'_{i,t-1})\alpha_4 + (N_i \times Z'_{i,t-1})\alpha_5 + e_{it} \end{aligned} \quad (1)$$

where $WhiteSup_{it}$ is equal to 1 if officer i is assigned to a white supervisor in year t and 0 otherwise. $Z_{i,t-1}$ is a vector of lagged annual work performance measures: the number of arrests for violent crimes, arrests for property crimes, arrests for non-index crimes, complaints made against the officer, use-of-force filings, and rate of strong use of force. X_{it} is a vector of controls for other characteristics, such as officer race, sex, birth year, and tenure, as well as unit and year fixed effects. B_i is a binary indicator variable equal to 1 if the officer is Black, H_i if Hispanic, A_i if Asian, and N_i if Native American; White officers are the reference group.

Table 3 reports estimates for average lagged annual work performance measures on the likelihood of being assigned to a white supervisor. Column 1 reports the coefficients for white officers (α_1), while column 2 reports the coefficients for the differences between Black and white officers (α_2). Because officers are assigned to supervisors at the Department level, we use all patrol officers assigned to a supervisor rather than the analysis sample that is restricted to officers whose supervisor assignment lasted one year. We also exclude units in years where there were only white supervisors because for those unit-years, the probability of having a white supervisor is equal to one.³³ To adjust for multiple

³³There are only nine unit-years where this is true. There are no units that had only Black supervisors in any

hypothesis testing, we calculate sharpened False Discovery Rate (FDR) q-values (Anderson, 2008; Benjamini et al., 2006). These are reported in square brackets underneath the unadjusted standard errors in parentheses. Standard errors are clustered at the unit level.

Table 3: Supervisor Race and Lagged Officer Work Measures

Dependent Variable: Coefficients for:	Supervisor is White	
	White Officer (1)	Black-White Difference (2)
Violent-crime Arrests	0.00189 (0.00191) [1.000]	0.000994 (0.00208) [1.000]
Property-crime Arrests	0.00256 (0.00183) [1.000]	9.85e-06 (0.00277) [1.000]
Non Index-crime Arrests	-0.000331 (0.000515) [1.000]	-3.14e-05 (0.000774) [1.000]
Complaints	0.00203 (0.00718) [1.000]	0.00528 (0.0130) [1.000]
TRR Filings	-0.0103 (0.00524) [1.000]	0.00583 (0.00765) [1.000]
Strong Force Ratio	-0.00402 (0.0164) [1.000]	-0.0333 (0.0347) [1.000]
Observations		9,938
R-squared		0.109
p-value for joint F-test	0.122	0.118

Source: CPD analysis sample.

Notes: The table reports estimates for lagged annual work measures on the likelihood of being assigned to a white supervisor. Coefficients for white officers are reported in column 1 and coefficients for the Black-white difference are reported in column 2. Estimates also control for officer race, birth year, tenure, unit, and year. Non-index arrests include arrests for non-property and non-violent crimes. The sample excludes units in years where there were only white supervisors. Standard errors clustered by unit are in parentheses. Sharpened False Discovery Rate q-values to adjust for multiple hypothesis testing are in square brackets. p-values for a joint F-test for white officers are reported in column 1 and for Black officers in column 2. *** sharpened $q < 0.01$, ** sharpened $q < 0.05$, * sharpened $q < 0.1$

Overall, prior-year work performance does not appear to predict supervisor race for given year.

both white or Black officers. Column 1 says that white officers' prior-year performance does not predict assignment to a white supervisor. The small and not-statistically significant coefficients in column 2 say that Black officers' prior-year performance does not differentially predict assignment to a white supervisor relative to white officers. For example, one additional violent-crime arrest last year increases the likelihood of having a white supervisor this year by about 0.19 percentage points for both white and Black officers.

We also conduct a joint F-test to see if these work measures are jointly significant in predicting assignment to a white supervisor. The p-values of 0.122 and 0.118 for white officers and Black officers, respectively, suggest they are not.³⁴ Moreover, the R-squared is low at 0.109, with 65 percent of the variation coming from unit and year fixed effects. This suggests that work performance measures explain very little of the variation in supervisor assignment.

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. Again, we do not see any statistically significant differences between white and Black officers assigned to white supervisors. Of course, we cannot test whether all unobservable traits are similar between white and Black officers assigned to the same supervisor. However, this analysis along with what we know about the administrative processes at CPD (a strong reliance on seniority) suggest that supervisor assignments do not take work measures into account.

5 Results

5.1 Black-White Gap by Arrest Record

This section examines whether white supervisors are equally likely to nominate Black and white officers conditional on their work performance. As the vast majority of non-

³⁴The joint F-test for Black officers is conducted on a modified version of equation (1) where the reference group is Black officers instead of white officers.

inations are crime-related, we ask how an officer’s arrest record affects his supervisor’s likelihood of nomination and whether this likelihood differs for minority officers. We estimate the following model:

$$\begin{aligned}
Nom_{it} = & \beta_0 + \left(\sum_{c=2}^3 \mathbb{1}\{Arrests_{i,t-1} = c\} \times \beta_1^c \right) + \left(B_i \times \sum_{c=2}^3 \mathbb{1}\{Arrests_{i,t-1} = c\} \times \beta_2^c \right) \\
& + \left(H_i \times \sum_{c=2}^3 \mathbb{1}\{Arrests_{i,t-1} = c\} \times \beta_3^c \right) + \left(A_i \times \sum_{c=2}^3 \mathbb{1}\{Arrests_{i,t-1} = c\} \times \beta_4^c \right) \\
& + \left(N_i \times \sum_{c=2}^3 \mathbb{1}\{Arrests_{i,t-1} = c\} \times \beta_5^c \right) + X'_{it}\alpha + \tau_t + \varepsilon_{it}
\end{aligned} \tag{2}$$

where 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 grouped into three bins (denoted by c): zero to one arrests; two to three arrests; and four or more arrests. We lag arrests because nominations must be submitted within 45 days of an incident.³⁵ The reference category is zero to one arrests last month. B_i is a binary indicator variable equal to 1 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.

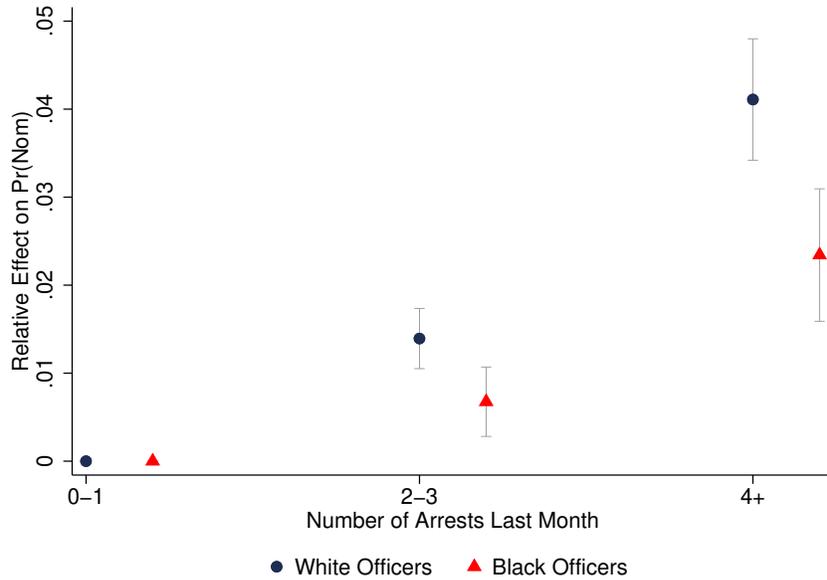
X_{it} is a vector of officer and supervisor characteristics. Officer controls include officer race, birth year, and tenure, as well as unit fixed effects. Supervisor controls are supervisor fixed effects. We also include fixed effects for year and month in τ_t . Standard errors are clustered at the supervisor level. We then calculate sharpened FDR q-values to adjust for multiple hypothesis testing.

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

³⁵The median number of days between the nomination and the incident date is 26 days.

number of arrests, then the elements of 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.

Figure 1: Nomination Likelihood by Lagged Arrests and Officer Race



Source: CPD analysis sample.

Notes: This figure depicts how the probability of nomination changes by the officer’s arrest record last month separately for white and Black officers. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year and tenure. Wings depict 95% confidence intervals using clustered standard errors.

Figure 1 depicts how the probability of nomination changes by an officer’s arrest record last month, separately for white and Black officers.³⁶ It highlights that there are increasing returns to having more arrests, with a marked increase for those with four or more arrests last month. Relative to having zero or one arrests last month, having four or more arrests increases the likelihood of a nomination by 4.1 percentage points ($p < 0.01$) for white officers. This is a three-fold increase from the baseline nomination likelihood of 1.3 percent.

³⁶Specifically, we plot estimates for β_1^c for white officers from equation (2). Then, we re-estimate equation (2) with Black officers as the reference group and plot the analogous estimates for β_1^c for Black officers.

The return to having more arrests, however, is less for Black officers compared to white officers; the return for Black officers is about 44 percent less ($= -0.018/0.041$).

In addition, the Black-white gap widens as the number of arrests increases and these differences are statistically significant (Table 4 column 1). The baseline Black-white nomination gap is -0.53 percentage points ($p < 0.01$). This gap widens by 1.7 percentage points ($p < 0.01$) for officers with four or more arrests. It is informative to interpret this disparity in the context of racial differences in work performance. For example, Black officers with four or more monthly arrests are at the 98th percentile of their distribution, while white officers are at the 92nd percentile of their distribution. Yet, white supervisors are 44 percent less likely to nominate Black officers relative to white officers.

As robustness checks, we also estimate a version of equation (2) with officer fixed effects and with supervisor and officer fixed effects. The patterns are highly similar across the different specifications.

In order to distinguish whether this behavior is due to in-group bias (favoritism) towards white officers or bias against Black officers, we examine whether white supervisors are also less likely to nominate Hispanic officers, another racial minority in the Chicago Police Department. Column 1 of Table 4 says that the Hispanic-white nomination gap is statistically and economically not significant.³⁷ However, white supervisors are 0.91 percentage points less likely to nominate Hispanic officers with four or more arrests relative to white officers with a similar number of arrests ($p < 0.05$). Relative to the Black-white disparity, however, the Hispanic-white gap is roughly half the size. These results suggest 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 officer makes around two arrests per month), but favor white officers when comparing officers with higher than average arrests.

In columns 2 and 3 of Table 4, we sharpen our focus on the type of arrest. Summary

³⁷This is true also for the Hispanic-white nomination gap among officers with zero or one arrest last month, which is 0.08 percentage points (p -value = 0.574).

Table 4: Impact of Lagged Arrests on Nomination Likelihood by Crime-Type

Arrests for:	Outcome Variable: Nominated		
	All Crimes (1)	Index Crimes (2)	Non-Index Crimes (3)
Two arrests last month	0.0139*** (0.00174) [0.001]	0.00997*** (0.00272) [0.001]	0.0157*** (0.00278) [0.001]
Three or more arrests last month	0.0411*** (0.00352) [0.001]	0.0188*** (0.00305) [0.001]	0.0425*** (0.00401) [0.001]
<i>Interactions with Black Officer</i>			
Two arrests last month	-0.00718*** (0.00253) [0.004]	-0.00695* (0.00382) [0.055]	-0.00811* (0.00475) [0.071]
Three or more arrests last month	-0.0177*** (0.00443) [0.001]	-0.0112*** (0.00386) [0.006]	-0.0163*** (0.00525) [0.003]
<i>Interactions with Hispanic Officer</i>			
Two arrests last month	0.000569 (0.00325) [0.168]	-0.000573 (0.00411) [0.422]	0.00116 (0.00475) [0.368]
Three or more arrests last month	-0.00908** (0.00506) [0.031]	-0.00280 (0.00464) [0.28]	-0.00686 (0.00570) [0.153]
Observations	177,144	177,144	177,144
Mean Pr(Nom) for Reference Group	0.013	0.026	0.016

Source: CPD analysis sample.

Notes: This table reports estimates for the impact of an officer's lagged arrest record on the probability of nomination by white supervisors. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year and tenure. Standard errors clustered by supervisor are in parentheses. Sharpened False Discovery Rate q -values to adjust for multiple hypothesis testing are in square brackets. *** sharpened $q < 0.01$, ** sharpened $q < 0.05$, * sharpened $q < 0.1$

statistics in Table 2 showed that white officers and Black officers make different types of arrests. If supervisors value certain types of arrests over others—particularly the type of arrests that white officers are making—this may explain the racial disparity in nominations. To address this concern, we estimate equation (2) separately for arrests for index crimes and for non-index crimes. The results are similar. Whether they made an arrest for a serious (index) crime or a less serious (non-index) crime, Black officers are still less likely to be nominated for an award relative to their white colleagues.

In Appendix Table A3, we examine Black supervisors' behavior as a complement to

white supervisors' behavior. This allows us to see whether both white and Black supervisors prefer white officers for awards (a la Arnold et al. (2018)) or if Black supervisors display in-group-favoritism. Although Black supervisors also are more likely to nominate officers with a greater number of arrests, they do not display statistically significant differences in their nomination likelihood between white, Black, and Hispanic officers.

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' (negative) beliefs about Black officers' work measures.

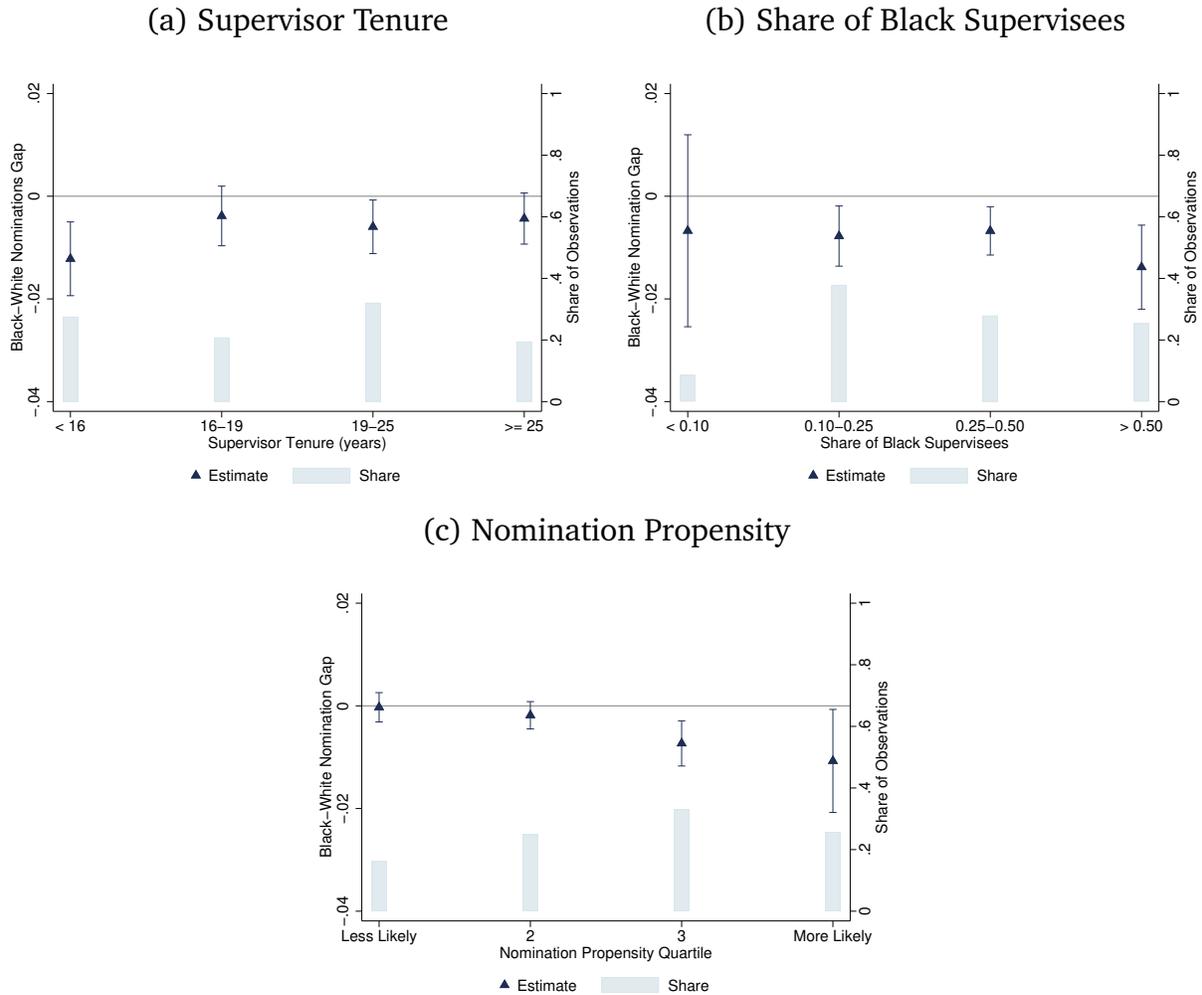
5.2 Black-White Gap by Supervisor Characteristics

We examine how the Black-white nomination gap changes by supervisor characteristics in Figure 2. 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 ($p < 0.01$).

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 most negative and statistically significant when the share of Black supervisees exceeds 50 percent. Among white supervisors with a majority of Black officers, the Black-white nomination gap is -1.4 percentage points ($p < 0.01$).

Panel C reports the Black-white nomination gap by the supervisor's propensity to nominate officers for an award. This is calculated as quartiles of a supervisor fixed effect from a regression of lagged officer arrests, officer tenure, officer fixed effects, the number of supervisees, and unit, month, and year fixed effects on nominations. Here, the Black-white gap is largest among those who are more prone to nominating officers at 1.1 percentage points ($p < 0.05$).

Figure 2: Black-White Nomination Gap by Supervisor Characteristics



Source: CPD analysis sample.

Notes: This figure depicts how the Black-white nomination gap changes by supervisor tenure (Panel A), by the share of Black supervisees (Panel B), and by nomination propensity (Panel C). All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year, tenure, lagged arrests, the number of supervisees, and the share of Black supervisees. Wings depict 95% confidence intervals using clustered standard errors.

5.3 Black-White Gap by Evaluation Quarter

We established that supervisors prefer to nominate white officers over Black officers conditional on the number of arrests. To interpret this disparity as a bias, one must assume that supervisors are equally aware of both white and Black officers' arrest activity. This may not be the reality in the Chicago Police Department. A 2016 report by the Po-

lice Accountability Task Force found little stability in supervisor-officer relationships. For example, 62 percent of officer shifts are not with their assigned supervisor.³⁸ As is the case in many workplaces, sergeants (supervisors) do not monitor officers (workers) constantly and thus may not be aware of their performance on every single shift. The information asymmetry from a lack of perfect monitoring may manifest in a Black-white gap in nominations if gathering information is costly (Bartoš et al., 2016).

To examine whether a learning gap can explain the nomination gap, we leverage an institutional feature that requires supervisors to conduct annual evaluations of their assigned officers. Although there may be little interaction between officers and supervisors on a daily basis—and, thus, why information acquisition is costly—we assume that supervisors are motivated to gather information and learn about an officer’s work record for the annual evaluation. Further, officer evaluations are randomly assigned across the calendar quarter, which allows us to disentangle any seasonal trends in nomination decisions from the effect of information gathering and to obtain a causal estimate of how knowledge about an officer’s work performance may influence a supervisor’s nomination likelihood.

This analysis can also provide some insight into the source of the racial disparity. If the Black-white nomination gap is a result of statistical discrimination, then we would expect the gap to shrink around the time of the evaluation as supervisors learn more about their officers’ work records. In contrast, if the gap were due to racial bias, then we would expect the gap to persist.

³⁸This statistic is calculated using shift assignments data between 2010 and 2019.

We estimate the following model:

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

where Nom_{it} is equal to 1 if officer i was nominated for an award in month t and equal to 0 if not. B_i is a binary indicator variable equal to 1 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 $RQ_{it} = 0$. The reference quarter is $RQ_{it} = -3$ or three quarters prior to the evaluation quarter of officer i .³⁹ The coefficients δ^q tell us how nomination likelihoods for white officers change across quarters. If information gathering 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 $RQ_{it} = -3$. If the Black-white difference does not change in subsequent quarters, then we expect β_1^q to be zero.

X_{it} is a vector of officer and supervisor characteristics. Officer controls include officer's race, birth year, and tenure, unit fixed effects, and the number of arrests the officer made last month. Supervisor controls are supervisor fixed effects. We also include fixed effects

³⁹As there are four possible calendar quarters in which an evaluation may take place, there are at most three calendar quarters prior to and at most three calendar quarters after the evaluation quarter. This means that the number of quarters relative to the evaluation can range from -3 to 3 for a total of seven relative quarters. For example, if an officer's evaluation is in December, then the relative quarters are Jan-Mar ($RQ = -3$), Apr-Jun ($RQ = -2$), Jul-Sep ($RQ = -1$), and Oct-Dec ($RQ = 0$). If, instead, an officer's evaluation is in January, the relative quarters are Jan-Mar ($RQ = 0$), Apr-Jun ($RQ = 1$), Jul-Sep ($RQ = 2$), and Oct-Dec ($RQ = 3$).

for year and month in τ_t . Standard errors are clustered at the supervisor level.

Table 5: Racial Difference in Nomination Likelihood by Quarter

Estimates for:	Outcome Variable: Nominated		
	White Officer (1)	Black-White Gap (2)	Hispanic-White Gap (3)
<i>Quarter relative to three quarters before evaluation</i>			
Two quarters pre-evaluation	0.0102*** (0.00318) [0.002]	-0.00838** (0.00390) [0.028]	-0.00659* (0.00491) [0.078]
One quarter pre-evaluation	0.0171*** (0.00327) [0.001]	-0.0128*** (0.00375) [0.002]	-0.00543 (0.00519) [0.117]
Evaluation quarter	0.0198*** (0.00374) [0.001]	-0.0139*** (0.00403) [0.002]	-0.00497 (0.00553) [0.136]
One quarter post-evaluation	0.0126*** (0.00387) [0.002]	-0.0141*** (0.00390) [0.001]	-0.00876** (0.00510) [0.049]
Two quarters post-evaluation	0.0104** (0.00419) [0.016]	-0.0133*** (0.00416) [0.002]	-0.00779* (0.00545) [0.071]
Three quarters post-evaluation	0.00622* (0.00486) [0.082]	-0.0125** (0.00520) [0.017]	-0.0140** (0.00598) [0.018]
Observations		171,094	
Mean Pr(Nom) for Reference Group		0.019	

Source: CPD analysis sample.

Notes: The table depicts how the quarterly probability of nomination changes relative to three quarters before the officer's evaluation. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year, tenure, and lagged arrests. Standard errors clustered by supervisor are in parentheses. Sharpened False Discovery Rate q-values to adjust for multiple hypothesis testing are in square brackets. *** sharpened $q < 0.01$, ** sharpened $q < 0.05$, * sharpened $q < 0.1$

Table 5 reports the coefficient estimates, clustered standard errors in parentheses, and sharpened FDR q-values in square brackets. Column 1 reports estimates for white officers (δ^q), column 2 reports estimates for the Black-white gap (β_1^q), and column 3 reports estimates for the Hispanic-white gap (β_2^q).

White officers are more likely to be nominated as they move closer to the evaluation quarter, but the increased likelihood falls afterward. Relative to three quarters before the evaluation, white officers are 1.7 percentage points more likely to be nominated in

the quarter before the evaluation and about 2 percentage points more likely to be nominated in the evaluation quarter (both $p < 0.01$). As white officers have a 1.9 percent chance of being nominated in the baseline quarter, this means the nomination likelihood essentially doubles in the evaluation quarter. One quarter post-evaluation, the relative likelihood drops but white officers are still 1.3 percentage points ($p < 0.01$) more likely to be nominated relative to the baseline quarter. This gradually tapers off and three quarters post-evaluation, white officers are 0.6 percentage points or 33 percent ($p < 0.1$) more likely to be nominated relative to the baseline group.

The baseline Black-white nomination gap is 0.39 percentage points and not statistically significant. In the quarters leading up to the evaluation, however, the gap widens and becomes statistically significant. It is the widest in the evaluation quarter and one quarter after the evaluation at 1.4 percentage points ($p < 0.01$) before shrinking.

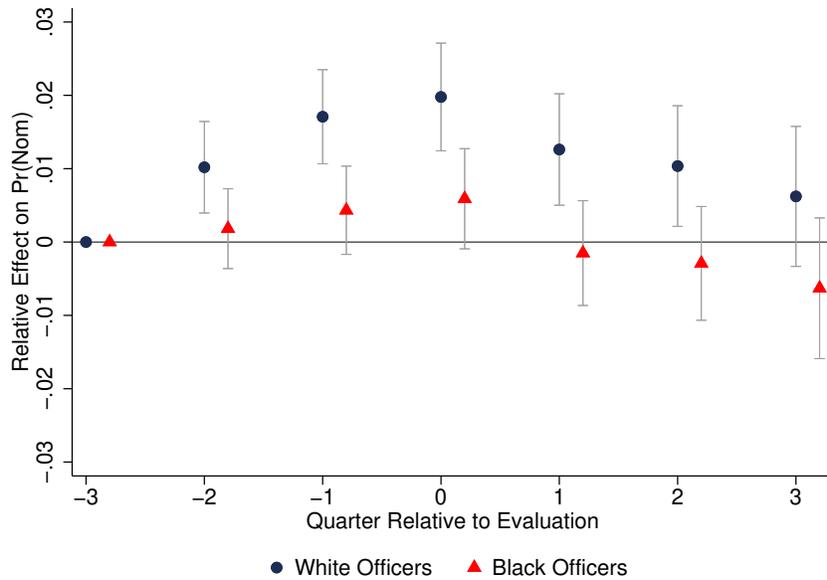
The nomination patterns for Hispanic officers, by contrast, are similar to those for white officers in the quarters leading up to and including the evaluation quarter. After the evaluation quarter, Hispanic officers are also less likely to be nominated, relative to white officers, but the drop in likelihood is about 59 to 62 percent less than the drop for Black officers, except for in the last quarter when the drop is slightly larger for Hispanic officers vs. Black officers.

In Appendix Table A4, we estimate a version of equation (3) with officer fixed effects and with supervisor and officer fixed effects as robustness checks. The patterns are similar across the different specifications.

Figure 3 provides some additional insight into the results. Is the Black-white gap widening because supervisors are acquiring more information about white officers in the evaluation quarter? Or because supervisors are gathering less information about Black officers? To answer this question, we plot the estimates for δ^q separately for white officers and Black officers.⁴⁰

⁴⁰Specifically, we plot estimates for δ^q when the reference group is white officers. Then, we re-estimate equation (3) with Black officers as the reference group and plot δ^q on the same graph.

Figure 3: Nomination Likelihood by Relative Quarter and Officer Race



Source: CPD analysis sample.

Notes: This figure depicts how the probability of nomination changes by quarter relative to three quarters before the officer’s evaluation, separately for white and Black officers assigned to white supervisors. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year, tenure, and lagged arrests. Wings depict 95% confidence intervals using clustered standard errors.

The inverted-U shape suggests that the nomination probability increases as we grow closer to the evaluation quarter then falls afterward. This pattern exists for both white and Black officers, but the increase is smaller for Black officers.

Although the positive estimates in Figure 3 indicate that Black officers are more likely to be nominated in the quarters leading up to their evaluation, the estimates are not statistically significant. In the quarter of evaluation, Black officers have a 0.6 percentage point increase in nomination likelihood ($p < 0.1$) relative to their baseline, which represents a 63 percent increase. After the evaluation, the estimates become trivially small and statistically not significant. Contrast this with the estimates for white officers, who are almost two percentage points (105 percent) more likely to be nominated in the evaluation quarter relative to their baseline and 1.3 percentage points (67 percent) more likely to be nominated in the quarter after the evaluation.

Another possible explanation is that officers may be endogenously reacting to the annual evaluation and changing their arrest behavior a la Glover et al. (2017). In Appendix Figure A1, we plot how arrests differ by relative quarter and officer race.⁴¹ Unlike for nomination likelihoods, the coefficients on relative quarters for arrests are stable and not statistically significant across relative evaluation quarters for white and Black officers.⁴² As such, it is unlikely that officers are endogenously changing their behavior around the time of their evaluation.

In Appendix Table A4, we examine Black supervisors' behavior. None of the point estimates are statistically significant, suggesting Black supervisors do not change their nomination behavior for white, Black, or Hispanic officers.

To summarize, white supervisors are more likely to nominate all officers in the evaluation quarter relative to the baseline quarter, suggesting that they are learning about or gathering information about their officers due to the annual evaluation. However, Black officers are consistently less likely to be nominated (15.5 to 42 percent less) suggesting that the learning is not translating into more nominations for Black officers and that white officers may choose to invest less time into learning about Black officers' performance. We further observe that white officers continue to benefit from the increased "attention" by receiving more nominations after the evaluation quarter, suggesting longer lasting bonds between white supervisors and white officers. By contrast, nominations for Black officers immediately revert to (and go below) their baseline nomination likelihood after the evaluation quarter. Our preferred interpretation of the results is that though a learning gap may exist, it cannot fully explain the nomination gap. Instead, we posit that the disparity can be explained by an advocacy gap: supervisors are less likely to choose to advocate for (nominate) Black officers relative to white officers, even after learning about their work performance.

⁴¹Specifically, we estimate a version of equation (3) with total arrests as the outcome variable.

⁴²Hispanic officers have positive coefficients for all relative quarters, but the coefficients are stable across quarters. In other words, there is no inverted-U shape as there is with nominations.

6 Experimental Evidence

The previous section presented evidence suggesting that a racial attention and advocacy gap may lead to a racial nomination gap. We explore this mechanism in an online experiment that seeks to examine whether evaluators choose to learn about minority officers.⁴³

We assume that there are two costs that CPD supervisors bear during a performance evaluation: a time cost of gathering and reviewing information about an officer's work record (e.g., logging into the Performance Evaluation system, asking other sergeants about their experience with the officer, etc.), and a psychic cost of engaging with the officer. The time cost does not differ by officer race, while the psychic cost may. If a supervisor exhibits racial bias, then we assume that the psychic cost is greater if the supervisor does not share the same race as the officer.⁴⁴ We ran an online experiment to isolate the psychic cost of engaging with people of different races. Specifically, we examine whether participants are less likely to engage with minorities when tasked with learning about their performance record.

Participants were asked to review real, anonymized CPD officer profiles and nominate one for an award.⁴⁵ Officer profiles display only demographic information, and participants must hover over the profile to reveal the officer's performance measures. We used mouse-tracking on the computer screen to measure which profiles participants hovered over—our primary measure of evaluator engagement. We also recorded the order in which participants moused over the profiles and the length of time participants engaged with a given profile. As an ancillary outcome, we measured which officer was ultimately nominated for an award.

⁴³The experiment was pre-registered in the AEA RCT Registry, AEARCTR-0005929.

⁴⁴For supervisors who do not exhibit racial bias, the psychic cost would be orthogonal to officer race.

⁴⁵This task was incentive compatible as participants were informed that their nomination decisions would be shared with CPD. After the experiment concluded, the researchers emailed aggregated nomination suggestions of the participants to the CPD. Further, because we used actual officer profiles, at no point during the experiment did we deceive participants or show unrealistic officer profiles.

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. Although CPD supervisors are a selected sample, we may not necessarily expect the two evaluator groups (police sergeants and online participants) to act very differently; Dickinson et al. (2015) finds that police commissioners are no different from non-police civilians when it comes to issuing rewards.⁴⁶

6.1 Experimental Design

The experiment had two main types of tasks. The first type examined whether evaluators were less likely to nominate minority officers. The second type examined whether evaluators were less likely to engage with (learn about) minority officers and whom evaluators ultimately nominated.

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. “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 this task, all of the officer’s demographic and work performance measures were visible to the evaluator.⁴⁷

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.⁴⁸ All officers were of “average quality”, defined as having zero or one civilian complaints and an average number of arrests. There were two itera-

⁴⁶In 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.

⁴⁷See Appendix Figure B2 for an example screenshot.

⁴⁸See Appendix Figure B3 for an example screenshot.

tions 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 types of 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 comparison).⁴⁹ For the second type of task, participants were restricted from uncovering any work performance measures for ten seconds and they were also not prevented from nominating an officer during the ten seconds. This was to ensure that participants had enough time to view 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 using Qualtrics surveys.⁵⁰ 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

⁴⁹These 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.

⁵⁰It 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.

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.

Demographically, MTurk evaluators are similar to CPD supervisors (Appendix Table B1). Both groups are majority white (64 percent vs 69.7 percent) and majority male, but CPD supervisors are more likely to be male than MTurk evaluators (80.8 percent vs 59.6 percent). MTurk evaluators are also slightly younger than the average CPD supervisor. The modal MTurk evaluatory is in the 26 to 35 age group, while the average CPD supervisor is 44 years old at the start of our analysis sample.

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

First, we seek to replicate the results from the CPD analysis and ask whether Black officers are 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) 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.

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

Table 6: Impact of being Black on Nomination Likelihood

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

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

= 0.111). There is no statistically significant Black-white nomination gap among Black evaluators. These results are consistent with the CPD analysis.

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 percentage-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 and, thus, this task passes the manipulation test.

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 learn about 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.

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)
<i>Panel A: Outcome Variable: Ever Moused Over</i>					
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)
<i>Panel B: Outcome Variable: First Mouseover</i>					
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)
<i>Panel C: Outcome Variable: Mouseover Duration (seconds)</i>					
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 evaluator engagement: 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.

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, 81.5 percent of Black officer profiles, 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.

When the officer pool is racially heterogeneous—that is, white, Black, and Hispanic officers are all represented—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 whom 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)
<i>Panel A: Outcome Variable: Moused Over</i>			
Black Officer	-0.0289 (0.0177)	-0.0707* (0.0386)	-0.0271 (0.0230)
Hispanic Officer	-0.0252 (0.0179)	0.0332 (0.0329)	-0.0475** (0.0237)
Observations	2,992	1,492	1,500
Mean Outcome for White Officer	0.842	0.838	0.852
<i>Panel B: Outcome Variable: First Mouseover</i>			
Black Officer	0.00355 (0.0210)	0.0399 (0.0409)	-0.0285 (0.0293)
Hispanic Officer	-0.0136 (0.0212)	0.0681 (0.0424)	-0.0578** (0.0293)
Observations	2,488	1,245	1,243
Mean Outcome for White Officer	0.300	0.289	0.328
<i>Panel C: Outcome Variable: Mouseover Duration (seconds)</i>			
Black Officer	0.431*** (0.107)	0.984*** (0.230)	0.347** (0.135)
Hispanic Officer	0.570*** (0.107)	0.697*** (0.224)	0.575*** (0.137)
Observations	2,488	1,245	1,243
Mean Pr(Nom) 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 evaluator engagement: 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

An alternative explanation to racial bias is that evaluators may be rationally acting on their beliefs about the relative productivity levels of Black and white officers. For example, if evaluators believe (whether correctly or not) that white officers are more productive than Black officers, then they would rationally choose to learn more about white officers because with limited attention, learning more about the relatively more-productive group will be more informative in a “cherry-picking market” where only the top candidate from a pool of candidates is nominated (Bartoš et al., 2016).

We attempt to distinguish between these two mechanisms (racial bias or rational behavior due to one’s beliefs) by examining which officer the evaluator eventually nominates after uncovering (learning) the officer profiles. It is worth noting that all officer profiles in this analysis are of average quality, so the decision to nominate an officer should solely reflect the evaluator’s preferences. More specifically, because all officer profiles are of a similar quality, there should be no reason to expect a Black-white nomination gap conditional on the evaluator having moused over both a Black and a white officer profile—if the reason for the initial disparity were statistical discrimination. On the other hand, if the reason were due to racial bias, then one would still expect to see a Black-white disparity in nomination likelihood after the evaluator learns that all officers are of the same quality.

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 ($p < 0.01$) more likely to be nominated (Panel A). This suggests that mouseover activity is a good measure of the evaluator’s interest in the officer.⁵¹ Workers who engage with Black or Hispanic officers are even more likely to nominate these minority officers—by an additional 8.7 percentage points ($p < 0.05$) and 14 percentage points ($p < 0.01$), respectively. Column 2 reveals that these results are driven by the predominantly white officer pool. Among profiles that were not moused over, Black and Hispanic officers are less likely to be nominated (6.3 percentage points ($p < 0.05$) and 15.7 percentage points ($p < 0.01$), respectively).

We also analyze first mouseover as a proxy for the intensity of an evaluator’s interest in an officer. None of the estimates are statistically significant, but it may be informative to discuss the signs and magnitudes of the coefficients. Conditional on being moused over, being moused over first increases the probability of nomination by 2.5 percentage points

⁵¹It was possible for participants to nominate an officer without mousing over their profile.

(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. Column 2 indicates that this finding is driven by a more racially homogeneous officer pool. A potential explanation for this may be that a minority officer will stand out in a predominantly white officer pool. Therefore, 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).

Among officers who were moused over, but were not the first mouseover, Black officers are 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. Although the point estimates are not statistically significant, the signs of the coefficients are consistent with our results using CPD administrative data (Section 5.3), which found that evaluators may not be updating their information about Black officers.

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 about the same as an additional second spent on a Black officer's profile. When the officer pool is predominantly white, however, the coefficient on the interaction between mouseover duration and Black officers is 0.039 ($p < 0.05$), 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 and learning about 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. There is some evidence that learning on the part of the evaluator may reduce the Black-white disparity; in a predominantly white officer pool, Black officers are more likely to be nominated the longer an evaluator spends reviewing the Black officer’s profile compared to a white officer’s profile.

It is informative to compare estimates when the officer pool is predominantly white vs. more racially diverse (Column 2 vs. Column 3). 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 ever moused over disappears (from 0.364 to -0.00966). This suggests that having a diverse police force may not actually eliminate the Black-white recognition gap.

Table 9: Impact of Officer Race on Nomination Likelihood

Officer Pool:	Outcome: Nominated Officer		
	All (1)	Predom. White (2)	Het. Race (3)
<i>Panel A: Ever Moused Over</i>			
Ever Moused Over	0.0882*** (0.0244) [0.001]	0.0785*** (0.0284) [0.007]	0.107 (0.0495) [0.184]
Black Officer	-0.0629** (0.0388) [0.044]	-0.380*** (0.0771) [0.001]	-0.294 (0.251) [0.318]
Hispanic Officer	-0.157*** (0.0378) [0.001]	-0.0102 (0.0684) [0.214]	-0.0177 (0.145) [0.672]
Ever Moused Over × Black Officer	0.0865** (0.0443) [0.027]	0.364*** (0.0797) [0.001]	-0.00966 (0.0649) [0.672]
Ever Moused Over × Hispanic Officer	0.140*** (0.0412) [0.002]	0.173** (0.0812) [0.017]	0.104 (0.0613) [0.220]
Observations	2,992	1,492	1,500
<i>Panel B: First Mouseover</i>			
First Mouseover	0.0251 (0.0259) [0.790]	-0.00466 (0.0313) [0.215]	0.109 (0.0483) [0.143]
Black Officer	-0.00694 (0.0268) [1.000]	-0.138 (0.0834) [0.138]	-0.205 (0.302) [0.698]
Hispanic Officer	-0.0330 (0.0326) [0.790]	0.11 (0.0600) [0.138]	0.0977 (0.163) [0.698]
First Mouseover × Black Officer	0.0922 (0.0493) [0.439]	0.194 (0.106) [0.138]	-0.0102 (0.0669) [1.000]
First Mouseover × Hispanic Officer	0.0242 (0.0528) [1.000]	0.172 (0.102) [0.138]	-0.113 (0.0701) [0.273]
Observations	2,488	1,245	1,243
<i>Panel C: Mouseover Duration (seconds)</i>			
Mouseover Duration	0.0442*** (0.00638) [0.001]	0.0397*** (0.00763) [0.001]	0.0544*** (0.0116) [0.001]
Black Officer	-0.0191 (0.0258) [0.225]	-0.198** (0.0726) [0.013]	-0.347 (0.229) [0.319]
Hispanic Officer	-0.105*** (0.0303) [0.003]	0.0863* (0.0582) [0.075]	0.0550 (0.128) [0.432]
Mouseover Duration × Black Officer	0.0102 (0.00966) [0.170]	0.0387** (0.0173) [0.026]	-0.0127 (0.0150) [0.432]
Mouseover Duration × Hispanic Officer	0.0260*** (0.00929) [0.006]	0.0140 (0.0150) [0.164]	0.0188 (0.0141) [0.319]
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 evaluator engagement 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. Robust standard errors are in parentheses. Sharpened False Discovery Rate q -values to adjust for multiple hypothesis testing are in square brackets. *** sharpened $q < 0.01$, ** sharpened $q < 0.05$, * sharpened $q < 0.1$

7 Conclusion

This paper studies whether a Black-white gap in award nominations may arise in part because white CPD supervisors do not advocate for their Black and white officers equally. First, we find that white supervisors are less likely to nominate Black officers relative to white officers with similar arrest records. We further find that the Black-white nomination gap widens with the number of arrests the officer made, which is the opposite of what one would expect if the disparity were due to a supervisor's (negative) beliefs about Black officers' work performance (i.e., statistical discrimination).

To provide stronger evidence on the underlying mechanism, we examine how the Black-white nomination gap changes as the supervisor (presumably) learns more information about his officers due to the annual evaluation. White supervisors are more likely to nominate all of their officers in the evaluation quarter relative to the baseline quarter, suggesting that supervisors are indeed learning new information. However, the persistent Black-white nomination gap suggests that learning additional information does not translate into more nominations for Black officers as it does for white officers. Collectively, these results suggest that supervisors are less likely to advocate for (nominate) Black officers relative to white officers, even after learning about their work performance. Again, this behavior is not consistent with statistical discrimination.

A nascent literature provides empirical evidence that discrimination can be reversed (Ayalew et al., 2021; Bohren et al., 2019; Lowe, 2019; Mousa, 2020).⁵² A critical component for the reversal is actual engagement with the minority group or a direct confrontation of the inaccurate belief (Hanna et al., 2014; Levine et al., 2021). Therefore, a crucial question for addressing racial bias in the workplace is whether managers pay attention to and advocate for their white and Black employees equally. Our paper provides suggestive evidence that managers do not, and thus, a racial attention and advocacy gap may lead to

⁵²The extent of this reversal may be limited to the intervention setting. For example, Mousa (2020) randomly assigned Iraq Christians to either an all-Christian soccer team or a team mixed with Muslims and found that treated players changed their behavior on the field but not off the field.

a racial gap in career progression.

This has important policy implications for policing. For example, diversity initiatives may be constrained by the extent to which officer bias carries over to their colleagues. In addition, 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 internal racial bias due to its effect on career advancement.

References

- Aigner, D. J. and G. G. Cain (1977). Statistical theories of discrimination in labor markets. *Ilr Review* 30(2), 175–187.
- 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.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103(484), 1481–1495.
- 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. 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.
- Arrow, K. J. (1973). The theory of discrimination. s. 3–33 in: O. ashenfelter/a. rees (hrsg.), *discrimination in labor markets*.
- Ayalew, S., S. Manian, and K. Sheth (2021). Discrimination from below: Experimental evidence from ethiopia. *Journal of Development Economics* 151, 102653.
- 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.
- Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3), 491–507.
- Bertrand, M. and S. Mullainathan (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review* 94(4), 991–1013.
- Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review* 109(10), 3395–3436.
- 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).

- Cullen, Z. B. and R. Perez-Truglia (2020). The old boys' club: Schmoozing and the gender gap. Working paper, National Bureau of Economic Research.
- 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.
- Egan, M. L., G. Matvos, and A. Seru (2020). When Harry fired Sally: The Double Standard in Punishing Misconduct. *Journal of Political Economy* Forthcoming.
- 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.
- Giuliano, L., D. I. Levine, and J. Leonard (2011). Racial Bias in the Manager-Employee Relationship An Analysis of Quits, Dismissals, and Promotions at a Large Retail Firm. *Journal of Human Resources* 46(1), 26–52.
- 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.
- Hanna, R., S. Mullainathan, and J. Schwartzstein (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics* 129(3), 1311–1353.
- 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.
- Lamont, M. (2018). Addressing recognition gaps: Destigmatization and the reduction of inequality. *American Sociological Review* 83(3), 419–444.

- Levine, S. S., E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, and D. Stark (2014). Ethnic diversity deflates price bubbles. *Proceedings of the National Academy of Sciences* 111(52), 18524–18529.
- Levine, S. S., C. Reypens, and D. Stark (2021). Racial Attention Deficit. 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.
- McKinsey & Company (February 21, 2021). Race in the workplace: The Black experience in the US private sector. Report.
- 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.
- 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.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review* 62(4), 659–661.
- Police Accountability Task Force (April 2016). Recommendations for reform: Restoring trust between the chicago police and the communities they serve. Technical report.
- Rim, N., B. Ba, and R. Rivera (2020). Disparities in Police Award Nominations: Evidence from Chicago. *AEA Paper and Proceedings* 110, 447–451.
- Sarsons, H. (2019). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. Working paper.
- Sarsons, H. (2021). Gender Differences in Recognition for Group Work. *Journal of Political Economy* 129(1), 101–147.

- U.S. Department of Justice (January 13, 2017). Investigation of the Chicago Police Department. Technical report.
- 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.

Online Appendix for
“The Black-White Recognition Gap in
Award Nominations”

A Additional Tables and Figures

Table A1: Department Awards

	Award Name and Description	Share
1	<i>Honorable Mention Certificate</i> for demonstrating outstanding performance above and beyond that required by the member's assignment.	87.48%
2	<i>Department Commendation</i> for an outstanding act or achievement that brings great credit to the Department and involves performance above and beyond that required by the member's basic assignment.	6.88%
3	<i>Unit Meritorious Performance Award</i> for exhibiting exceptional professional skill and conduct during a coordinated action.	1.24%
4	<i>Special Commendation</i> for making a significant impact on public safety or crime prevention.	0.97%
5	<i>Problem Solving Award</i> for an exemplary effort to identify, analyze, and successfully respond to causes, conditions, and problems that may lead to crime and neighborhood disorder.	0.79%
6	<i>Joint Operations Award</i> for efforts and participation in a broad multi-agency joint operation/event, spanning several days or more, significantly contributing to the overall successes of the operation.	0.77%
7	<i>Lifesaving Award</i> for a successful effort in saving a human life that involved exceptional courage or performance.	0.65%
8	<i>Police Officer of the Month</i> for performance of duty during a specific month was characterized by such exceptional professional skill that it merits recognition by the entire Department.	0.32%
9	<i>Traffic Stop of the Month Award</i> for excellence in conducting professional traffic stops that result in quality arrests.	0.26%
10	<i>Superintendent's Award of Valor</i> for an act of outstanding bravery or heroism by which the member has demonstrated in great degree the characteristics of selflessness, personal courage, and devotion to duty.	0.24%
11	<i>Arnold Mireles Special Partnership Award</i> for making a significant impact upon the quality of life within their community by identifying and resolving problems.	0.16%
12	<i>Superintendent's Award of Merit</i> for an outstanding accomplishment that has resulted in improved administration, improved operations, or substantial savings in manpower or operational costs, wherein the member has gone far beyond the requirements of their normal assignment.	0.14%
13	<i>Top Gun Arrest Award</i> for exceptional commitment to the recovery of illegal firearms.	0.06%
14	<i>Police Blue Star Award</i> is granted to any sworn member who has been seriously, critically, or fatally injured while in the performance of police duty.	0.03%
15	<i>Police Blue Shield Award</i> is granted to any sworn member who, as a result of accidental causes, has been seriously, critically, or fatally injured while in the performance of police duty.	0.01%
16	<i>Chicago Police Leadership Award</i> for exemplary service, dedication, and leadership.	0.01%
17	<i>Special Service Award</i> for contributing to any event that has a significant impact upon the historical direction and operations of the Department.	0.01%
18	<i>Superintendent's Award of Tactical Excellence</i> for exceptional tactical skills or verbal approaches and techniques to mitigate any deadly force situation resulting in the saving or sustaining of a human life.	0%

Source: Chicago Police Department Special Order S01-01-01 "Description and Eligibility for Department Awards", retrieved from <http://directives.chicagopolice.org/directives/> and CPD analysis sample.

Notes: This table lists the 18 department awards considered in the analysis in decreasing order of prevalence.

Table A2: Evaluation Quarter and Due Dates by Start Month

Quarter	Anniversary Date Month of the Member	The Quarter the Member Will Be Evaluated	Due Date of the Evaluation
1st	January, February, March	4th	30 January
2nd	April, May, June	1st	30 April
3rd	July, August, September	2nd	30 July
4th	October, November, December	3rd	30 October

Source: Chicago Police Department, Career Development Directive, Employee Resource E05-01, Section IX, B. Retrieved from <http://directives.chicagopolice.org/directives/data/a7a56e3d-12887ea9-ce512-887e-c3dce7cd73e28d57.html?ownapi=1>

Table A3: Impact of Lagged Arrests on Nomination Likelihood: Robustness Checks

Outcome Variable: Nominated			
Supervisor Race:	White (1)	White (2)	Black (3)
Two to three arrests last month	0.00702*** (0.00175) [0.001]	0.00676*** (0.00176) [0.001]	0.00774 (0.00390) [0.14]
Four or more arrests last month	0.0201*** (0.00248) [0.001]	0.0170*** (0.00230) [0.001]	0.0290*** (0.00741) [0.001]
<i>Interactions with Black Officer</i>			
Two to three arrests last month	-0.00392* (0.00247) [0.093]	-0.00377 (0.00248) [0.107]	-0.00373 (0.00493) [0.37]
Four or more arrests last month	-0.00946** (0.00364) [0.013]	-0.00716* (0.00351) [0.06]	-0.0122 (0.0100) [0.235]
<i>Interactions with Hispanic Officer</i>			
Two to three arrests last month	0.00258 (0.00313) [0.205]	0.00313 (0.00310) [0.232]	0.00200 (0.00961) [0.517]
Four or more arrests last month	-0.00210 (0.00372) [0.236]	-0.00220 (0.00359) [0.238]	-0.0165 (0.00981) [0.173]
Mean Pr(Nom) for Reference Group	0.013	0.013	0.006
Observations	177,144	177,144	30,588
Controls for:			
Officer	Yes	Yes	
Supervisor		Yes	Yes

Source: CPD analysis sample.

Notes: Each column is a separate regression. This table reports estimates for the impact of an officer's lagged arrest record on the probability of nomination. All estimates include unit, month, and year fixed effects, and control for officer birth year and tenure. Standard errors clustered by supervisor are in parentheses. Sharpened False Discovery Rate q -values to adjust for multiple hypothesis testing are in square brackets. *** sharpened $q < 0.01$, ** sharpened $q < 0.05$, * sharpened $q < 0.1$

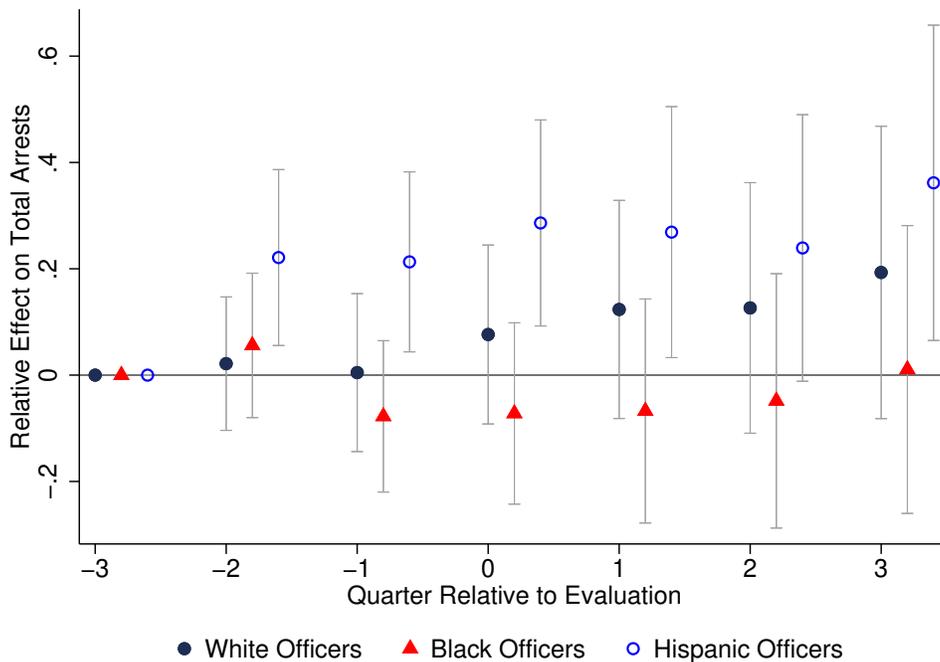
Table A4: Racial Difference in Nomination Likelihood by Quarter: Robustness Checks

Supervisor Race:	Outcome Variable: Nominated		
	White (1)	White (2)	Black (3)
<i>Quarter relative to two quarters before evaluation</i>			
Two quarters pre-evaluation	0.0104** (0.00378)	0.0103** (0.00379)	0.018 (0.00836)
One quarter pre-evaluation	0.0168*** (0.00481)	0.0164** (0.00482)	0.0175 (0.00769)
Evaluation quarter	0.0191*** (0.00595)	0.0185** (0.00595)	0.0191 (0.00804)
One quarter post-evaluation	0.0117 (0.00722)	0.0110 (0.00722)	0.00473 (0.00750)
Two quarters post-evaluation	0.00842 (0.00875)	0.00753 (0.00876)	0.0112 (0.0108)
Three quarters post-evaluation	0.00378 (0.00972)	0.00279 (0.00974)	0.00253 (0.00870)
<i>Two quarters pre-evaluation ×</i>			
Black Officer	-0.00665 (0.00431)	-0.00660 (0.00432)	-0.0168 (0.00795)
Hispanic Officer	-0.00692 (0.00531)	-0.00736 (0.00532)	-0.0223 (-0.0108)
<i>One quarter pre-evaluation ×</i>			
Black Officer	-0.0108** (0.00440)	-0.0106* (0.00440)	-0.0123 (0.00749)
Hispanic Officer	-0.00329 (0.00560)	-0.00347 (0.00560)	-0.00416 (0.0118)
<i>Evaluation quarter ×</i>			
Black Officer	-0.0113** (0.00464)	-0.0112* (0.00465)	-0.0146 (0.00791)
Hispanic Officer	-0.00170 (0.00615)	-0.00185 (0.00616)	0.000237 (0.0129)
<i>One quarter post-evaluation ×</i>			
Black Officer	-0.0107* (0.00480)	-0.0105* (0.00481)	-0.00365 (0.00691)
Hispanic Officer	-0.00479 (0.00606)	-0.00497 (0.00606)	0.00166 (0.0104)
<i>Two quarters post-evaluation ×</i>			
Black Officer	-0.00966 (0.00528)	-0.00941 (0.00530)	-0.0132 (0.00990)
Hispanic Officer	-0.00120 (0.00685)	-0.00140 (0.00685)	-0.00107 (0.0105)
<i>Three quarters post-evaluation ×</i>			
Black Officer	-0.00675 (0.00642)	-0.00648 (0.00643)	-0.00925 (0.00764)
Hispanic Officer	-0.00611	-0.00617	-0.00839
Mean Pr(Nom) for Reference Group	0.019	0.019	.012
Observations	171,094	171,094	29,413
Controls for:			
Officer	Yes	Yes	
Supervisor		Yes	Yes

Source: CPD analysis sample.

Notes: Each column is a separate regression. The table depicts how the quarterly probability of nomination changes relative to three quarters before the officer's evaluation. All estimates include unit, month, and year fixed effects, and control for officer tenure and lagged arrests. Standard errors clustered by supervisor are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure A1: Number of Arrests by Relative Quarter and Officer Race



Source: CPD analysis sample.

Notes: This figure depicts how the number of total arrests by quarter relative to three quarters before the officer's evaluation, separately for white, black, and Hispanic officers. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year and tenure. Wings depict 95% confidence intervals using clustered standard errors.

B Online Experiment

The experiment was conducted on Amazon Mechanical Turk (MTurk) in July 2020. It was pre-registered in the AEA RCT Registry, AEARCTR-0005929. We recruited 411 MTurk workers (hereafter “workers”) who were compensated \$1.20 for completing a survey experiment. Table B1 reports summary statistics on all 411 workers. Figure B1 plots the distribution of workers’ states of residence.

We included three data quality checks to identify bots and to ensure workers paid attention during the survey. For the analysis, we decided to include workers who passed at least two of the three data quality checks. This restriction reduces our final analysis sample to 407 workers.

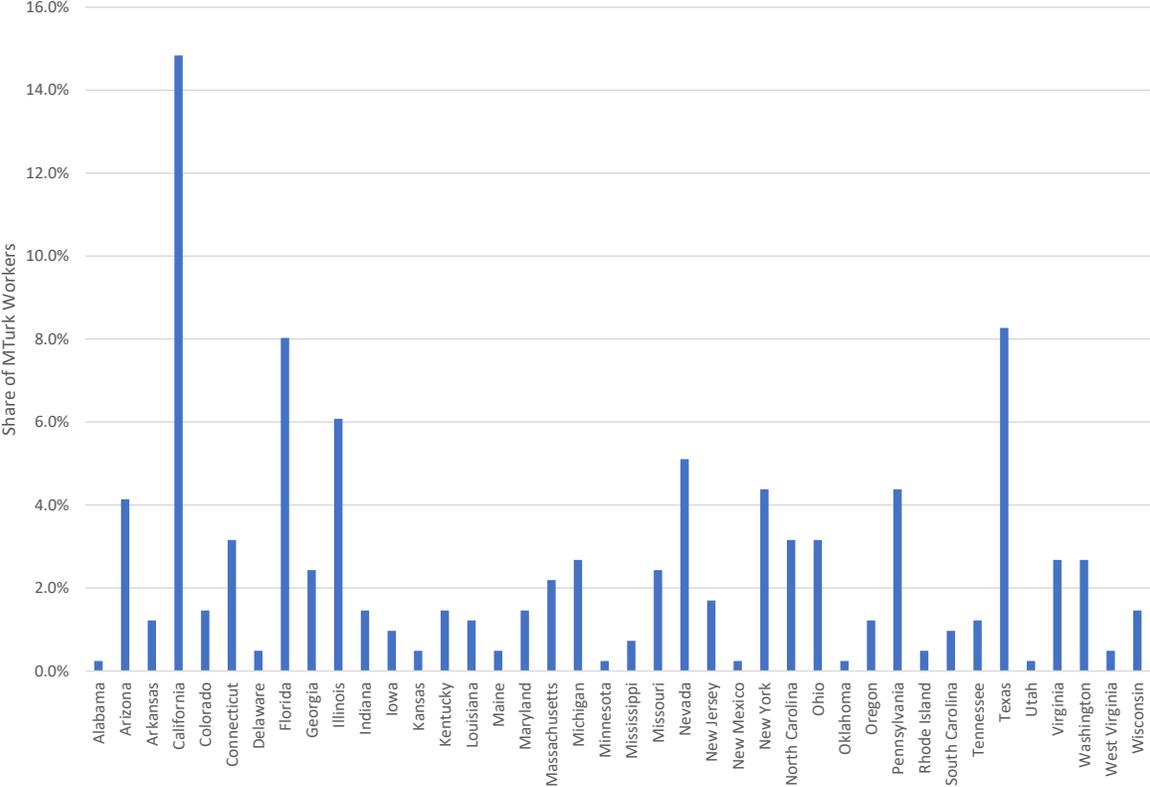
To avoid deception in our survey, we used real officer profiles but used officer initials to preserve officers’ identities. Workers were informed that the profiles belonged to real officers but were not told which agency they were from. Further, we informed workers that their nominations would be relayed to the police department. This was to achieve incentive compatibility. After the experiment ended, the Chicago Police Department was informed of survey results.

Table B1: Summary Statistics

	N	%
Race		
Black	48	11.7%
Hispanic	66	16.1%
White	263	64.0%
Other	21	5.1%
Prefer not to answer	10	2.4%
Missing	3	0.7%
Female	166	40.4%
Age		
18-25	53	12.9%
26-35	189	46.0%
36-45	78	19.0%
46-55	55	13.4%
56+	35	8.5%
Missing	1	0.2%
Is English your first language?		
Yes	401	97.6%
No	5	1.2%
Missing	5	1.2%
Length of Residency in US		
< 1 yr	6	1.5%
More than 1 yr but less than 3 yrs	21	5.1%
More than 3 yrs but less than 6 yrs	16	3.9%
More than 6 yrs	365	88.8%
Missing	3	0.7%
Number of Surveys (MTurk Workers)	411	

Source: MTurk survey data.

Figure B1: Distribution of MTurk Worker State of Residence



Source: MTurk survey data.

Figure B2: Screenshot of Pairwise Comparison Task

06

Which of these two officers would you recommend for an award?

<input type="radio"/>	Initials	A.L.
	Race	White
	Sex	Male
	Age	51
	Experience	9.33
	Total arrests	24
	Civilian complaints	1

<input type="radio"/>	Initials	R.N.
	Race	Black
	Sex	Male
	Age	47
	Experience	8.08
	Total arrests	35
	Civilian complaints	0

Figure B3: Screenshot of Group Comparison Task

Here are four officer profiles. Select the one you would recommend for an award. Once the black boxes appear, you will have 30 seconds to make your decision. The boxes will turn red 5 seconds before your time is up.

<input type="radio"/>	Initials	K.B.
	Race	White
	Sex	Male
	Age	28
	Experience	
	Total arrests	
	Civilian complaints	

<input type="radio"/>	Initials	D.S.
	Race	Black
	Sex	Male
	Age	38
	Experience	
	Total arrests	
	Civilian complaints	

<input type="radio"/>	Initials	S.D.
	Race	Hispanic
	Sex	Male
	Age	32
	Experience	
	Total arrests	
	Civilian complaints	

<input type="radio"/>	Initials	S.O.
	Race	White
	Sex	Male
	Age	41
	Experience	
	Total arrests	
	Civilian complaints	