

# The Black-White Recognition Gap in Award Nominations\*

Nayoung Rim<sup>†</sup>

Roman Rivera<sup>‡</sup>

Andrea Kiss<sup>§</sup>

Bocar Ba<sup>¶</sup>

May 1, 2022

There is substantial evidence showing racial bias in firms' hiring decisions, but less is known about bias in career progression. We construct a dataset from the second largest US police department to estimate the Black-white gap in award nominations. Leveraging institutional features to obtain plausibly causal estimates, we find that white supervisors are less likely to nominate Black officers over other officers conditional on their work performance. This appears to be driven by supervisor bias in advocacy decisions, rather than statistical discrimination. Given the reliance on subjective evaluations for promotions in many organizations, our findings 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, advocacy, promotions, supervisors, police, learning

---

\*We would like to thank Peter Arcidiacono, Pat Bayer, Sandra Black, Kerwin Charles, Rob Garlick, Rachel Greenspan, Kareem Haggag, Damon Jones, Ilyana Kuziemko, 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.

<sup>†</sup>U.S. Naval Academy, Department of Economics. Email: [rim@usna.edu](mailto:rim@usna.edu)

<sup>‡</sup>Columbia University, Department of Economics. Email: [r.g.rivera@columbia.edu](mailto:r.g.rivera@columbia.edu)

<sup>§</sup>Carnegie Mellon University, Department of Social and Decision Sciences. Email: [akiss@andrew.cmu.edu](mailto:akiss@andrew.cmu.edu)

<sup>¶</sup>Duke University, Department of Economics. Email: [bocar.ba@duke.edu](mailto: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,<sup>1</sup> 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,<sup>2</sup> 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.<sup>3</sup>

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

---

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

<sup>2</sup>According 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/>

<sup>3</sup>For 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.

the CPD, awards are used in important career decisions, such as merit promotions.<sup>4</sup> 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.<sup>5</sup>

To protect against endogenously formed supervisor-supervisee relationships, we leverage the fact that 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.

We first establish that white supervisors are 28 to 40 percent less likely to nominate Black officers relative to other officers, conditional on the number and type of arrest.<sup>6</sup> The negative 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).<sup>7</sup>

To explore this further, we ask whether the racial nomination gap exists because white supervisors are choosing to learn about white officers over Black officers. 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—

---

<sup>4</sup>Merit 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.

<sup>5</sup>We 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.

<sup>6</sup>Award nominations are typically related to a reported crime, which is why we condition on the number of arrests. For 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.

<sup>7</sup>For 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).

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.

To examine whether a racial learning gap can explain the racial nomination 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.<sup>8</sup> This analysis provides 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 Black officers’ work records.

We find that white supervisors are more likely to nominate both white and Black officers in the evaluation quarter relative to the baseline quarter, suggesting that the racial nomination gap at least partially reflects statistical discrimination. However, the negative Black-white nomination gap remains constant across all quarters, suggesting that white supervisors are not choosing to advocate for white and Black officers equally. For example, there is a steady increase in nomination likelihood in the quarters leading up to the evaluation for white officers, which is not present for Black officers. Additionally, white officers experience a sustained increase in nomination likelihood after the evaluation that Black officers do not experience.

---

<sup>8</sup>The 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 further evidence of a racial advocacy gap when we leverage a natural discontinuity built into a specific award and find that white supervisors are discontinuously more likely to nominate white officers on the threshold of qualifying for this award but not their Black officers. To provide stronger evidence on this mechanism, we conduct 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 performance 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.

Although we cannot definitively determine the source of discrimination due to a lack of data on beliefs and preferences, our collective findings suggest that supervisor bias in advocacy decisions, rather than statistical discrimination, is driving the racial nomination gap. The implications from our findings are significant because awards help Black officers disproportionately more than white officers when it comes to career advancement and merit promotions. Yet, Black officers are less likely to be nominated for and win department awards. Thus, our findings underscore the potential power of awards in reducing racial disparities—conditional on a minority officer having advocates in the workplace.

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).<sup>9</sup> 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 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.<sup>10</sup> 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 an issue between the police and the public, but also within departments and, thus, that simply hiring minority officers may be limited in its efficacy.

## 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.<sup>11</sup> 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

---

<sup>9</sup>Hengel (2019) and Sarsons (2019) find that female minorities are less likely to be acknowledged for their work, while the others focus on racial minorities.

<sup>10</sup>See, for example, Ajilore and Shirey (2017); Antonovics and Knight (2009); Anwar and Fang (2006); Bacher-Hicks and de la Campa (2020); Close and Mason (2006); 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).

<sup>11</sup>Between 2012 and 2013, three districts were dissolved leaving 22 geographic districts.

(Police Accountability Task Force, 2016).<sup>12</sup>

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, defined as a unique combination of 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.<sup>13</sup>

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.<sup>14</sup> To clarify, a *supervisor* by our definition differs from the *sergeant on duty*. 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 their off-days are rotated 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.<sup>15</sup> 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

---

<sup>12</sup>Additionally, officers may bid for district transfers when vacancies occur. Successful bidders are selected based on seniority.

<sup>13</sup>Employee Resource E05-05, Section III.A., available at <http://directives.chicagopolice.org> and CPD Sergeant Written Assessment Study Briefing 2013, Appendix A, available at [https://www.chicago.gov/content/dam/city/depts/dhr/general/CPD\\_Sergeant\\_Assessment\\_Study\\_Briefing\\_2013.pdf](https://www.chicago.gov/content/dam/city/depts/dhr/general/CPD_Sergeant_Assessment_Study_Briefing_2013.pdf).

<sup>14</sup>According to shift-assignment data, 62 percent of officer shifts between 2010 and 2019 were not with the officer's annually assigned supervisor.

<sup>15</sup>A duty cycle is typically four on-duty days followed by two off-duty days.

report found this system “prevents supervisors from establishing mentoring relationships 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 evaluations and merit promotions.<sup>16</sup>

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

---

<sup>16</sup>For information on performance evaluations, see Chicago Police Department, Career Development Directive, Employee Resource E05-01, Section IV.H., available at <http://directives.chicagopolice.org>. For information on merit promotions, see Section III.E.2, Employee Resource E05-05, available at <http://directives.chicagopolice.org>.



number of times an officer may be nominated, as long as the nominations are for different incidents.<sup>17</sup> 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 through a nomination from a high-ranking member of the CPD command staff, even if their examination score did not place them at the top of the list. Merit promotions rely on a variety of officer qualifications, such as awards. For example, a one-standard-deviation increase in the number of awards is correlated with a 4.3 percentage-point increase ( $p < 0.01$ ) in receiving a merit promotion, conditional on passing the sergeant exam (Appendix Table A1). Award nominations are also an important predictor; a one-

---

<sup>17</sup>There 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>).

standard-deviation increase in the number of award nominations is correlated with a 3.9 percentage-point increase ( $p < 0.01$ ) in receiving a merit promotion.

## 3 Data

This section describes the 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.<sup>18</sup> 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

---

<sup>18</sup>Not 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.

paper, the term “supervisor” refers to a sergeant who is officially assigned to conduct a police officer’s annual evaluation in a given calendar year.

*Awards* The awards dataset provides information on all department award nominations between 2004 and 2017. The dataset includes the award name, the individual being nominated, the requester, request date, and the final status of the nomination (approved, deleted, or denied).<sup>19</sup>

*Arrests* The arrests dataset contains information on all arrests made by police officers. The dataset includes detailed information on 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.<sup>20</sup> 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.<sup>21</sup> Non-index crimes capture crimes that are less serious such as municipal code violations, traffic violations, and drug abuse violations.<sup>22</sup>

*Complaints* The complaints dataset contains all recorded allegations of misconduct filed against an officer between 2000 and 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

---

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

<sup>20</sup>In 96.9 percent of cases, the release date is on the same day or the day after the arrest date, and 100 percent of release dates are within four days of the arrest.

<sup>21</sup>Violent index crimes are murder, aggravated assault/battery, robbery, and rape. Property index crimes are larceny theft, burglary, arson, and motor vehicle theft.

<sup>22</sup>A comprehensive list of crime categories can be found at [http://gis.chicagopolice.org/clearmap\\_crime\\_sums/crime\\_types.html](http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html).

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 harm on or injure the subject.<sup>23</sup>

*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 *Police Officer* rank who are always assigned to a geographic district and officer-supervisor assignments 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.<sup>24</sup> Appendix Table A2 provides a description of these awards. The vast majority of awards are Honorable Mention Certificate Awards, which are relatively broad in scope, compared to, say, the Police Blue Star Award for an officer injury. They also 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. In Section 3.2, we will examine which work measures best predict an award nomination.

## 3.2 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

<sup>23</sup>Strong use of force may or may not necessarily involve weapons. For example, take-downs, kicks, and punches are also strong uses of force.

<sup>24</sup>Most awards are open to all Department members. One example of an exception is the Thomas Wortham IV Military and Community Service Award, which is awarded to current or former members of the U.S. Armed Services.

in our sample joined the force in 2000 at age 30. This indicates that at the start of our analysis dataset (year 2009), the average officer had been on the force for nine years.

Relative to police officers, the racial makeup of supervisors 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 to 15 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.

Notes: "Police Officers" are entry-level officers. "Supervisors" are sergeants who are officially assigned to conduct a police officer's annual evaluation.

Table 2 presents summary statistics on various work measures. The first row is the probability of being nominated for an award in a particular month. For example, the average officer has a 2.5 percent chance of being nominated in a given month, which equates to about a 30 percent chance of being nominated in a given year. White and Hispanic officers have slightly higher than average likelihoods at 3 percent and 3.2 percent, respectively, while the likelihood for Black officers is half the sample average (1.3 percent). The Black-white difference is statistically significant at the 1 percent level. The second

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)
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.

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 ( $p < 0.01$ ).<sup>25</sup>

In Table 3, we examine which factors best predict award nominations. Column 1 reports the Pearson’s correlation coefficient between various work measures and the number of nominations. Arrests has the strongest correlation at 0.162. Columns 2 through 8 report estimates from a regression of lagged monthly work measures on award nominations.

The top row reports the adjusted R-squared of the regression, and the rows below report coefficient estimates and standard errors for officer work measures. Unit and time fixed effects explain very little of the variation in nominations ( $\bar{R}^2 = 0.004$ ). Including use of force and complaints marginally improves the fit to 0.005. When we include arrests, however, the adjusted R-squared increases five-fold to 0.028.

In the most comprehensive specification (column 8), the number of arrests has the largest weight and is statistically significant at 1 percent. Officer tenure is also statistically significant at 1 percent, but the magnitude is trivial. Officer complaints is similar in magnitude to arrests but weakly significant at 10 percent. Further, the coefficient on complaints is positive, suggesting that nominations are likely made based on superficial knowledge of an officer’s arrest activity rather than a thorough review of the incident. In summary, these results facilitate our empirical strategy of conditioning on arrests when estimating the Black-white nomination gap.

---

<sup>25</sup>Although the data reveal a racial 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). Studies show that increasing the number of Black and female officers can reduce crime victimization and increase reports of domestic violence (Harvey and Mattia, 2020; Miller and Segal, 2018). These outcome measures, which are important measures of social welfare, are not captured by arrests nor would they appear on an officer’s record. Relatedly, others have shown that Black officers make fewer discretionary stops and use force less often than their white colleagues (Ba et al., 2021; Hoekstra and Sloan, 2020). For this reason, we do not necessarily interpret differences in overall arrests as differences in policing quality. At the same time, however, arrests are the strongest predictor of award nominations.

Table 3: Associations between Work Measures and Award Nominations

	Correlations	Outcome Variable: Number of Award Nominations						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Adjusted R-squared</b>		<b>0.001</b>	<b>0.004</b>	<b>0.005</b>	<b>0.005</b>	<b>0.005</b>	<b>0.028</b>	<b>0.028</b>
TRR Filings	0.0341			0.0410*** (0.00492)	0.0297*** (0.00742)	0.0268*** (0.00725)	-0.0109 (0.00691)	-0.0118* (0.00692)
Strong Force Ratio	0.0316				0.0152* (0.00810)	0.0148* (0.00811)	0.0103 (0.00807)	0.00995 (0.00805)
Complaints	0.029					0.0346*** (0.00711)	0.0116* (0.00659)	0.0114* (0.00660)
Total Arrests	0.1619						0.0158*** (0.000954)	0.0154*** (0.000972)
Officer Tenure	-0.064							-0.000805*** (0.000150)
Observations	242,185	250,872	250,872	242,185	242,185	242,185	242,185	242,185
Year FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unit FE			Yes	Yes	Yes	Yes	Yes	Yes

Source: CPD analysis sample.

Notes: This table reports estimates for lagged monthly work measures on the number of award nominations received by the officer's supervisor. All work measures except for tenure and unit are lagged by one month. Standard errors clustered by officer are in parentheses. \*\*\*  $p < 0.01$ ,

\*\*  $p < 0.05$ , \*  $p < 0.1$



## 4 Identifying Assumptions

This section outlines the empirical strategy to examine whether supervisors are less likely to nominate Black officers conditional on their work performance and whether supervisors are less likely to learn about Black officers. We exploit two institutional features of the Chicago Police Department 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 avoid the problem of endogeneously formed relationships and to approximate random assignment of an officer's race to a supervisor.<sup>26</sup> Although the vast majority of supervisor assignments last one year (78 percent), we may be concerned that some officer-supervisor assignments may have been arranged outside of the dispassionate assignment system. Since our outcome variable is supervisor nominations, this would result in omitted variable bias if, for example, certain white officers were more likely to bypass the dispassionate assignment system and chose their supervisors. Therefore, we restrict our analysis sample to all supervisor-officer assignments that last one year in order to minimize the number of endogenously formed supervisor relationships.<sup>27</sup> 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 nomination 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 quar-

---

<sup>26</sup>About 96 percent of officers are assigned to a supervisor in January of each calendar year.

<sup>27</sup>In Appendix Tables A3 and A4, we redo our analysis to include all supervisor assignments instead of restricting the sample to those that lasted one year. We find that officers in a longer-term supervisor relationship are more likely to be nominated, lending support to our hypothesis that these relationships may have been endogeneously formed. For example, the mean nomination likelihood for white officers in longer-term relationships is 4.6 percent compared to 1.9 percent for those in one-year relationships. Further, the Black-white nomination gap widens when we include these longer-term relationships. As white officers are more likely than Black officers to have a longer-term supervisor relationship, the larger Black-white nomination gap is likely due to endogenously formed relationships between white officers and their supervisors. Thus, this robustness check supports our decision to restrict the analysis sample to officer-supervisor relationships that last exactly one year.

ter 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.<sup>28</sup> In Appendix Figure A1, we test whether the evaluation quarter is equally distributed across officer race. The p-value on the Chi-square test is 0.516, indicating that we cannot reject the null hypothesis that there is no difference in the distribution of evaluation quarter across the difference races.

To estimate the causal impact of information gathering, it is important for the evaluation period to differ across officers particularly because new supervisors are assigned at the same time every year. 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 the gap may 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 Black 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

---

<sup>28</sup>Specifically, all CPD candidates who passed the written exam 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. Appendix Table A5 lists the evaluation quarters and evaluation due dates by start month. For example, if an officer started the police academy in July (Q3), then his annual evaluation must take place in the second quarter of every calendar year (Q2).

may appear to be bias against Black officers by white supervisors, but in reality it would be the result of officers sorting based on work performance measures and race. 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 4 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 ( $\alpha_1$ ), while column 2 reports the coefficients for the differences between Black and white officers ( $\alpha_2$ ). 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.<sup>29</sup> We

<sup>29</sup>As a robustness check, we test our identifying assumption when the sample is restricted to supervisor assignments that lasted one year (Appendix Table A6). The results are similar. The coefficient on non-index-crime arrests is statistically significant, but the significance goes away when we correct for multiple

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.<sup>30</sup> To adjust for multiple hypothesis testing, we calculate sharpened False Discovery Rate (FDR) q-values (Anderson, 2008; Benjamini et al., 2006).<sup>31</sup> These are reported in square brackets underneath the unadjusted standard errors in parentheses. Standard errors are clustered at the unit level.

---

hypothesis testing. Most importantly, the joint F-tests do not find statistical significance.

<sup>30</sup>There are only nine unit-years where this is true. There are no units that had only Black supervisors in any given year.

<sup>31</sup>A concern with multiple hypothesis testing is that the rate of false rejections increases mechanically as the number of null hypotheses increases. For example, in Table 4, there are 12 null hypotheses. If we test these hypotheses one by one, then the probability of one or more false rejections when using a 5-percent critical value is almost 46 percent ( $= 1 - 0.95^{12}$ ). Therefore, we use the sharpened q-values to reduce the likelihood of false rejections and avoid reporting false associations. Importantly, the sharpened q-values do not change our conclusions.

Table 4: Supervisor Race and Lagged Officer Work Measures

Coefficients for:	Outcome Variable: Supervisor is White	
	White Officer (1)	Black-White Diff. (2)
Violent-crime Arrests	0.00183 (0.00149) [1.000]	0.000639 (0.00181) [1.000]
Property-crime Arrests	0.00245 (0.00177) [1.000]	0.000327 (0.00206) [1.000]
Non Index-crime Arrests	-0.000151 (0.000408) [1.000]	0.000393 (0.000741) [1.000]
Complaints	9.55e-05 (0.00539) [1.000]	0.00275 (0.00753) [1.000]
TRR Filings	-0.00973* (0.00566) [1.000]	0.0109 (0.00850) [1.000]
Strong Force Ratio	0.00750 (0.0115) [1.000]	-0.0113 (0.0234) [1.000]
Observations		31,263
R-squared		0.091
p-value for joint F-test	0.294	0.535

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.

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

Overall, prior-year work performance does not appear to predict supervisor race for both white or Black officers. Column 1 says that white officers' prior-year performance does not predict assignment to a white supervisor.<sup>32</sup> The small and not-statistically-

<sup>32</sup>The coefficient on TRR filings is statistically significant at the 10 percent level, but the statistical significance goes away when we correct for multiple hypothesis testing. More importantly, the magnitude is

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 0.18 percentage points for both white and Black officers, and this estimate is not statistically significant.

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 0.294 and 0.535 for white officers and Black officers, respectively, suggest they are not.<sup>33</sup> Moreover, the R-squared is low at 0.091, with 78 percent of the variation coming from unit and year fixed effects. This suggests that work 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 (i.e., 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 award nominations are tied to arrests, we ask how an officer's arrest record affects his supervisor's likelihood of nomination and whether this likelihood differs for Black officers. We estimate the following

---

trivial (-0.00973) and the Black-white difference in TRR filings is not statistically significant.

<sup>33</sup>The 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.

model:

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

where  $Nom_{ijt}$  is equal to 1 if officer  $i$  was nominated for an award by supervisor  $j$  in month  $t$  and equal to 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 by one month because nominations must be submitted within 45 days of an incident.<sup>34</sup> 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.

$X_{ijt}$  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  $\tau_t$ . 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  $\beta_1^c$ , which tell us how the nomination likelihood changes as the number of arrests last month increases, and  $\beta_2^c$ , which tell us how the Black-white nomination difference changes by the number of arrests. We expect  $\beta_1^c$  to be positive and increasing in the number of arrests. If the Black-white gap in award nominations does not vary by the number of arrests, then the elements of  $\beta_2^c$  will be zero. As the baseline Black-white gap is negative, a negative  $\beta_2^c$  indicates that the Black-white gap widens with

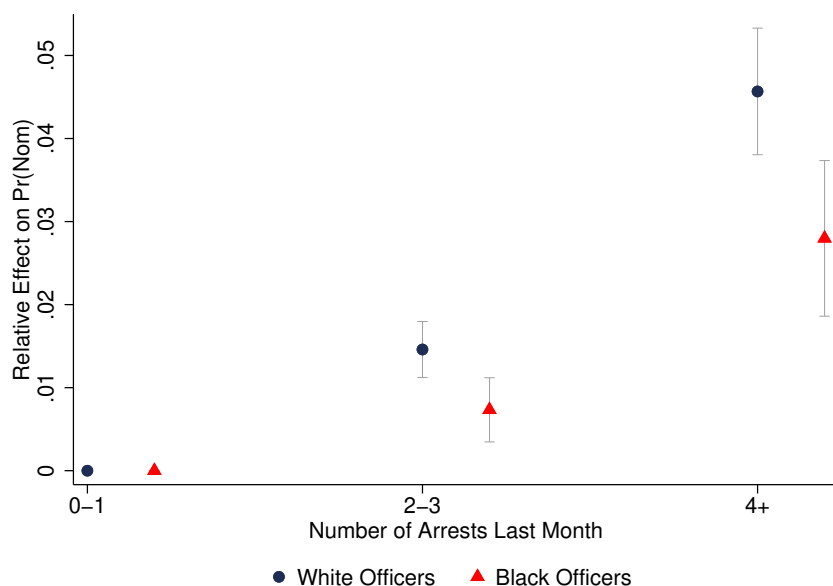
---

<sup>34</sup>The median number of days between the nomination and the incident date is 26 days.

the number of arrests, whereas a positive  $\beta_2^c$  indicates that the Black-white gap narrows with the number of arrests.

Figure 1 depicts how the probability of nomination changes by an officer's arrest record last month, separately for white and Black officers.<sup>35</sup> 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 arrest last month, having four or more arrests increases the likelihood of a nomination by 4.6 percentage points ( $p < 0.01$ ) for white officers. This is a three-fold increase from the baseline nomination likelihood of 1.3 percent. The return to having more arrests, however, is about 39 percent less ( $= -0.0177/0.0457$ ) less for Black officers compared to white officers, and this is statistically significant at the 1 percent level.

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.

<sup>35</sup>Specifically, we plot estimates for  $\beta_1^c$  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  $\beta_1^c$  for Black officers.



Table 5 reports estimates for the Black-white nomination difference by number of arrests ( $\beta_2^c$ ). The Black-white gap among officers with zero to one arrest is -0.54 percentage points ( $p < 0.01$ ). This gap widens by 1.8 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 39 percent less likely to nominate Black officers relative to white officers.

In columns 2 and 3 of Table 5, we sharpen our focus on the type of arrest. Summary 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 (more serious) crimes and for non-index (less serious) crimes. We find a large, negative, and statistically significant Black-white nomination gap for both index-crime arrests and non-index-crime arrests. The existence of a racial disparity among arrests for more serious crimes, like violent and property crimes, is consistent with our argument that the Black-white nomination gap reflects a racial advocacy gap rather than solely being a result of Black officers vs. white officers making different types of arrests.

We conduct several robustness checks where we include additional measures of officer performance, such as complaints and use of force (Appendix Table A7), and officer fixed effects and both officer and supervisor fixed effects (Appendix Table A8). The patterns are highly similar across the different specifications. We also estimate the Black-white nomination gap by predicted nomination probability, and the takeaway is very similar to our main results: the Black-white nomination gap widens as the predicted nomination probability increases (Appendix Figure A2).<sup>36</sup>

<sup>36</sup>Specifically, we use an out-of-sample group to predict a probability of nomination for each in-sample officer. Then, we estimate the Black-white nomination gap by quantiles of predicted probability.

Table 5: 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 to three arrests last month	0.0146*** (0.00171) [0.001]	0.0122*** (0.00244) [0.001]	0.0201*** (0.00259) [0.001]
Four or more arrests last month	0.0457*** (0.00388) [0.001]	0.0286*** (0.00601) [0.001]	0.0571*** (0.00576) [0.001]
<i>Interactions with Black Officer</i>			
Two to three arrests last month	-0.00727*** (0.00247) [0.003]	-0.00598* (0.00336) [0.061]	-0.00873** (0.00423) [0.055]
Four or more arrests last month	-0.0177*** (0.00530) [0.002]	-0.0256*** (0.00709) [0.001]	-0.0160* (0.00881) [0.076]
<i>Interactions with Hispanic Officer</i>			
Two to three arrests last month	0.000838 (0.00319) [0.153]	-0.00249 (0.00377) [0.205]	0.00134 (0.00441) [0.149]
Four or more arrests last month	-0.0121** (0.00552) [0.012]	-0.00695 (0.00884) [0.205]	-0.0134* (0.00832) [0.085]
Observations	176,552	176,552	176,552
Mean Pr(Nom) for Reference Group	0.013	0.027	0.016
Black-White Nomination Gap in Reference Group	-0.005	-0.010	-0.006

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

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 5 says that the Hispanic-white nomination gap among officers with two or three arrests is statistically and economically not significant.<sup>37</sup> However, white supervisors are 1.21 percentage points less likely to nominate Hispanic

<sup>37</sup>This is true also for the Hispanic-white nomination gap among officers with zero or one arrest last month, which is 0.11 percentage points (p-value = 0.433).

officers with four or more arrests relative to white officers with a similar number of arrests ( $p < 0.05$ ). Taken together, the 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 Appendix Table A9, we examine Black and Hispanic supervisors' behavior as a complement to white supervisors' behavior. In contrast to white supervisors, we do not find statistically significant estimates for the Black-white nomination gap among Black and Hispanic supervisors, indicating in-group favoritism by white supervisors. Additionally, the point estimates are 31 to 54 percent smaller for minority supervisors compared to white supervisors.<sup>38</sup> Under the homophily assumption that Black supervisors may be less prejudiced against Black officers than white supervisors, these results lend suggestive evidence that the racial nomination gap among white supervisors may be due to supervisor bias.<sup>39</sup>

To summarize, we find that Black officers are nominated less frequently than white officers even conditional on the number and type of arrest. Specifically, given a white officer and a Black officer, both of whom made zero or one arrest for an index crime last month, the white officer is 1 percentage point more likely than the Black officer to be nominated for an award. This is an effect size of 37.6 percent. If we look at officers with four or more arrests for index crimes, the Black-white nomination gap widens by 2.6 percentage points,

---

<sup>38</sup>One exception is the coefficient on the Black-white nomination gap among officers with 2-3 arrests by Hispanic supervisors, which is positive.

<sup>39</sup>We would also like to note that the negative signs on the Black officer interactions for Black supervisors do not necessarily imply that white supervisors have no bias. For example, if racial bias does influence promotion decisions, then the set of Black officers who are promoted to sergeant and become supervisors may be intrinsically different from the average Black police officer and may instead be more similar intrinsically to the average white officer. In this scenario, minorities in white-dominated firms may be selected into leadership roles partly on their willingness to be team players who act to protect the status quo, which may result in unsupportive behavior towards fellow minority colleagues. This type of behavior has been called value threat (Srivastava and Sherman, 2015), stereotype threat (Spencer et al., 2016), and collective threat (Cohen and Garcia, 2005). Srivastava and Sherman (2015) find empirical support for this theory in regards to the gender wage gap using longitudinal personnel records from a U.S. information services firm. In regards to law enforcement, Van Maanen (1975) documents changes in the attitudes of police recruits throughout their early careers and finds evidence of a police socialization process that stresses a "don't rock the boat" approach to policing.

an effect size of 51.4 percent.<sup>40</sup> The widening gap with increasing arrests is the opposite of what one would expect if the disparity were due to white supervisors' (negative) beliefs about Black officers' work measures.

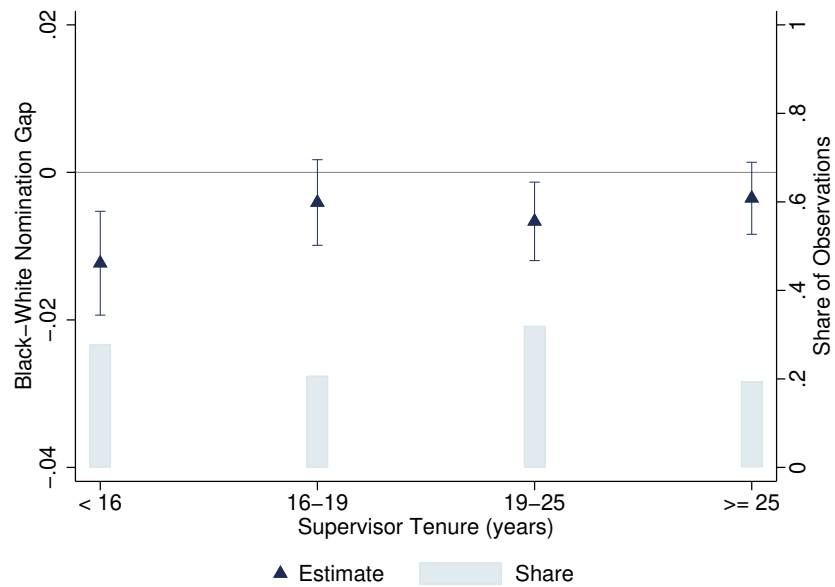
We also examine how the Black-white nomination gap changes by supervisor tenure, with quantiles defined by the 25th, 50th, and 75th percentile values (about 16 years, 19 years, and 25 years, respectively). Figure 2 shows that less experienced white supervisors have the largest negative Black-white nomination gap at -1.2 percentage points ( $p < 0.01$ ), and this estimate is statistically significantly different from the other tenure categories at the 1 percent level. The patterns are consistent with two potential explanations.

First, the nomination gap may be the result of in-group favoritism towards white officers and a racial advocacy gap, as the gap is largest among the newest supervisors who likely have stronger ties to their officers compared to the supervisors who have not been at entry-level rank for some time. At the same time, the shrinking of the racial nomination gap as supervisor tenure increases may also be explained by a learning story, as more experienced supervisors likely have a better sense of the distributions of officer performance by race. In the next section, we explore this further and examine whether a racial learning gap is driving the racial nomination gap.

---

<sup>40</sup>The mean nomination likelihood for white officers with four or more index-crime arrests is 6.9 percent, and the Black-white difference for officers in this group is -3.56 percentage points.

Figure 2: Black-White Nomination Gap by Supervisor Tenure



Source: CPD analysis sample.

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

## 5.2 Black-White Gap by Evaluation Quarter

In the previous section, we established that supervisors prefer to nominate white officers over Black officers conditional on the number and type of arrest. An implicit assumption in that analysis is that supervisors are equally aware of both white and Black officers' arrest activity. This may be not be the reality in the Chicago Police Department. A 2016 report by the Police Accountability Task Force found little stability in supervisor-officer relationships. For example, 62 percent of officer shifts are not with their assigned supervisor.<sup>41</sup> As is the case in many workplaces, sergeants (supervisors) do not monitor officers (workers) continuously 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

<sup>41</sup>This statistic is calculated using shift assignments data between 2010 and 2019.

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 decision.

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 Black officers’ work records.

We estimate the following model:

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

where  $Nom_{ijt}$  is equal to 1 if officer  $i$  was nominated for an award by supervisor  $j$  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  $EQ_{it} = 0$ . The baseline quarter is  $EQ_{it} = -3$

or three quarters prior to the evaluation quarter of officer  $i$ .<sup>42</sup> The coefficients  $\delta^q$  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,  $EQ = 0$ .

The third term in parentheses interacts the Black indicator variable and the relative-quarter indicator variables. The coefficients  $\beta_1^q$  depict how the Black-white nomination gap evolves relative to  $EQ_{it} = -3$ . If the relative Black-white difference does not change in subsequent quarters, then we expect  $\beta_1^q$  to be zero.

$X_{ijt}$  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 for year and month in  $\tau_t$ . Standard errors are clustered at the supervisor level.

Table 6 reports the coefficient estimates, clustered standard errors in parentheses, and sharpened FDR q-values in square brackets. Column 1 reports estimates for white officers ( $\delta^q$ ), column 2 reports estimates for the Black-white gap ( $\beta_1^q$ ), and column 3 reports estimates for the Hispanic-white gap ( $\beta_2^q$ ).

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 1.75 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. This gradually peters off but during the six months post-

---

<sup>42</sup>As 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 ( $EQ = -3$ ), Apr-Jun ( $EQ = -2$ ), Jul-Sep ( $EQ = -1$ ), and Oct-Dec ( $EQ = 0$ ). If, instead, an officer's evaluation is in January, the relative quarters are Jan-Mar ( $EQ = 0$ ), Apr-Jun ( $EQ = 1$ ), Jul-Sep ( $EQ = 2$ ), and Oct-Dec ( $EQ = 3$ ).

evaluation, white officers are still 1.1 to 1.3 percentage points ( $p < 0.01$ ) more likely to be nominated relative to the baseline quarter.

The Black-white nomination gap in the baseline quarter ( $EQ = -3$ ) is 0.4 percentage points and not statistically significant. In all subsequent quarters, however, the gap becomes negative and statistically significant. 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 Hispanic-white difference is about 38 to 65 percent smaller than the Black-white difference, except for in the last quarter when the difference is slightly larger (14 percent) for Hispanic officers vs. Black officers. Further, the estimates are mostly not statistically significant.

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  $\delta^q$  separately for white officers and Black officers.<sup>43</sup>

The hump 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. Further, although the positive estimates in Figure 3 suggest that Black officers are more likely to be nominated in the quarters leading up to their evaluation relative to their baseline, none of these estimates are statistically significant. In the quarter of evaluation, Black officers have a 0.7 percentage point increase in nomination likelihood ( $p < 0.05$ ) relative to their baseline mean of 0.0094, which represents a 74 percent increase. After the evaluation, however, the estimates become trivially small and lose statistical significance. Contrast this with the estimates for white officers, who are 2.1 percentage points (110 percent) more likely to be nominated in

<sup>43</sup>Specifically, we plot estimates for  $\delta^q$  when the reference group is white officers. Then, we re-estimate equation (3) with Black officers as the reference group and plot  $\delta^q$  on the same graph.



Table 6: 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	0.0116*** (0.00308) [0.001]	-0.00971** (0.00383) [0.012]	-0.00574 (0.00486) [0.098]
One quarter pre-evaluation	0.0175*** (0.00322) [0.001]	-0.0130*** (0.00372) [0.003]	-0.00479 (0.00512) [0.132]
Evaluation quarter	0.0207*** (0.00368) [0.001]	-0.0136*** (0.00399) [0.003]	-0.00480 (0.00543) [0.132]
One quarter post-evaluation	0.0133*** (0.00383) [0.003]	-0.0139*** (0.00392) [0.001]	-0.00864* (0.00500) [0.041]
Two quarters post-evaluation	0.0111*** (0.00415) [0.009]	-0.0130*** (0.00410) [0.003]	-0.00775 (0.00533) [0.066]
Three quarters post-evaluation	0.00689 (0.00481) [0.066]	-0.0123** (0.00521) [0.017]	-0.0140** (0.00591) [0.017]
Observations		176,552	
Mean Pr(Nom) for Reference Group		0.019	
Black-White Nomination Gap in Reference Quarter		0.004	

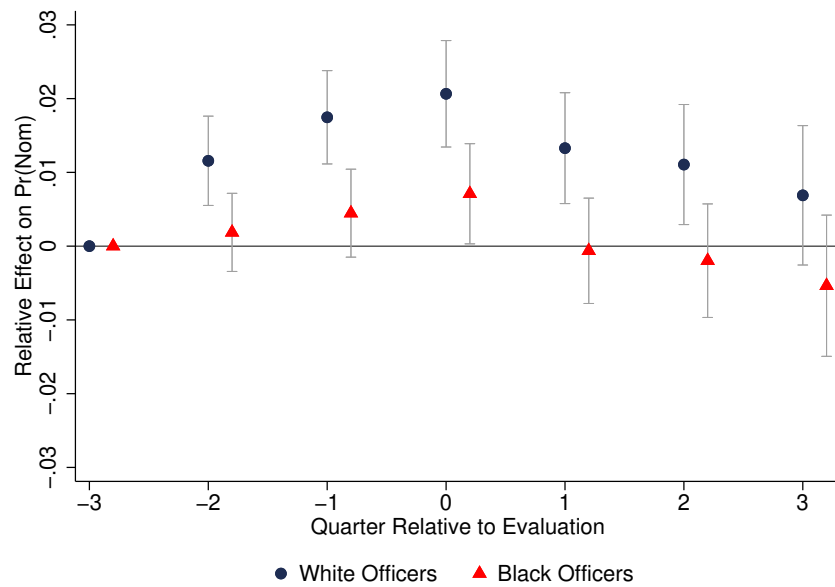
Source: CPD analysis sample.

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

the evaluation quarter relative to their baseline mean of 0.019 and 1.3 percentage points (70 percent) more likely to be nominated in the quarter after the evaluation.

Taken together, these results suggest that, although information gathering may be an important channel in a supervisor's nomination decision (because both white and Black officers experience an increase in nomination likelihood in the evaluation quarter), white supervisors are not choosing to advocate for white and Black officers equally. For example, there is a steady increase in nomination likelihood in the quarters leading up to the evaluation for white officers, which is not present for Black officers. Additionally, white officers experience a sustained increase in nomination likelihood after the evaluation that Black officers do not experience. Put another way, although Black officers are more likely

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.

to be nominated in the evaluation quarter relative to the baseline quarter, the Black-white nomination gap remains fairly constant across all quarters.<sup>44</sup>

In Appendix Table A10, we report results by supervisor race. Just like their white colleagues, Black and Hispanic supervisors are more likely to nominate their officers in the quarter before and the quarter of evaluation, indicating that minority supervisors are also learning new information about their officers. In contrast to white supervisors, however, the Black-white nomination gap for Black supervisors is not statistically significant over all the quarters, and the estimates are smaller than those for white supervisors.<sup>45</sup> Again, under the assumption that Black supervisors may be less prejudiced against Black officers than white supervisors, these results offer supportive evidence of supervisor bias among

<sup>44</sup>The estimates in column 2 are not statistically significantly different from each other.

<sup>45</sup>Estimates are 14 to 91 percent smaller with one being 18 percent larger.

white supervisors.

One potential alternative explanation to bias is that officers may be endogenously reacting to the annual evaluation and changing their arrest behavior a la Glover et al. (2017).<sup>46</sup> In Appendix Figure A3, we plot how arrests evolve by relative quarter and officer race.<sup>47</sup> We do not see any statistically significant changes in arrest activity by both white and Black officers. As such, it is unlikely that the increase in nomination likelihood for white officers is due to their endogenously increasing their arrest activity around the time of their evaluation.

To summarize, we find that white supervisors are more likely to nominate both white and Black officers in the evaluation quarter relative to the baseline quarter, suggesting that they are learning about or gathering additional information about their officers due to the annual evaluation. This suggests that the Black-white nomination gap at least partially reflects statistical discrimination. However, the negative Black-white nomination gap remains constant across all quarters. This is because white officers continue to receive more nominations after the evaluation quarter, suggesting longer lasting bonds between white supervisors and white officers, whereas 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 racial learning gap may exist, it cannot fully explain the racial nomination gap. Instead, we posit that the disparity can be explained by a racial advocacy gap: supervisors are less likely to advocate for (nominate) Black officers relative to white officers, even after learning about their work performance.

### 5.3 Black-White Gap by Past Awards

In this section, we further examine the evidence for a racial advocacy gap by leveraging a natural discontinuity built into a specific award, the Honorable Mention Ribbon award.

<sup>46</sup>Glover et al. (2017) find that minority workers lower their effort levels when assigned to biased managers.

<sup>47</sup>Specifically, we estimate a version of equation (3) with total arrests as the outcome variable.

This award is awarded to any officer who has “demonstrated outstanding performance and has received a minimum of 50 Honorable Mentions [Certificates].”<sup>48</sup>

To test whether white supervisors are more likely to advocate for white officers compared to Black officers, we first examine whether there is a discontinuous increase in the nomination probability for officers who are close to the 50-award threshold. We would expect this to be true if supervisors were indeed choosing to advocate for their officers. Then, we examine whether there is a Black-white nomination gap among these officers on the margin.

The regression model is:

$$\begin{aligned}
 Nom_{ijt} = & \beta_0 + \left( \sum_{c=2}^5 \mathbb{1}\{Awards_{i,t-1} = c\} \times \beta_1^c \right) + \left( B_i \times \sum_{c=2}^5 \mathbb{1}\{Awards_{i,t-1} = c\} \times \beta_2^c \right) \\
 & + \left( H_i \times \sum_{c=2}^5 \mathbb{1}\{Awards_{i,t-1} = c\} \times \beta_3^c \right) + \left( A_i \times \sum_{c=2}^3 \mathbb{1}\{Awards_{i,t-1} = c\} \times \beta_4^c \right) \\
 & + \left( N_i \times \sum_{c=2}^3 \mathbb{1}\{Awards_{i,t-1} = c\} \times \beta_5^c \right) + X'_{ijt}\alpha + \tau_t + \varepsilon_{ijt}
 \end{aligned} \tag{4}$$

where  $Nom_{ijt} = 1$  if officer  $i$  was nominated by supervisor  $j$  for an Honorable Mention Certificate award in month  $t$  and equal to 0 if not. We calculate the total number of Honorable Mention Certificates that each officer has won by the end of each month and classify officers into five categories by their lagged monthly cumulative number. Category 1 is officers with 0 to 13 total awards (50th percentile threshold), Category 2 is officers with 14 to 37 total awards (90th percentile threshold), Category 3 is officers with 38 to 44 total awards, Category 4 is officers with 45 to 49 total awards, and Category 5 is officers with 50 or more total awards. The marginal category is Category 4, whose officers are just below the 50-award threshold to qualify for the Honorable Mention Ribbon award.

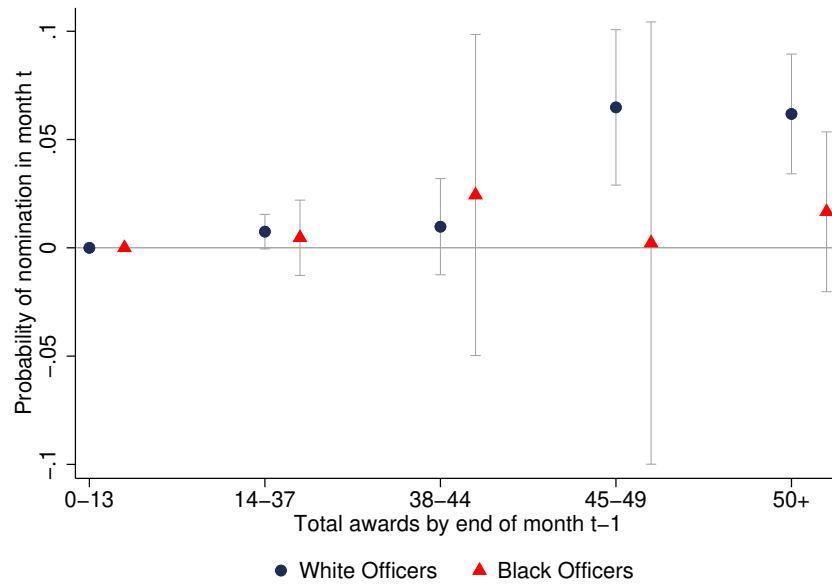
$B_i$  is a binary indicator variable equal to 1 if the officer is Black,  $H_i$  if Hispanic,  $A_i$  if

<sup>48</sup>Chicago Police Department, Special Order S01-01-01, available at <http://directives.chicagopolice.org>. The Honorable Mention Certificate makes up the bulk of all awards (see Appendix Table A2).

Asian, and  $N_i$  if Native American. White officers are the reference group.  $X_{ijt}$  is a vector of officer and supervisor characteristics. Officer controls include officer race, birth year, unit, lagged arrests, and tenure. Supervisor controls are supervisor fixed effects. We also include month and year fixed effects in  $\tau_t$ . Standard errors are clustered at the supervisor level.

Figure 4 reports estimates for  $\beta_1^c$ , which tell us how the nomination probability in month  $t$  changes depending on the officer's cumulative number of Honorable Mention Certificates, separately for white officers and for Black officers.

Figure 4: Impact of Past Awards on Nomination Likelihood



Source: CPD analysis sample.

Notes: This figure depicts how the number of past Honorable Mention Certificates affects the nomination likelihood by white supervisors this month, separately for white officers and Black officers. The marginal category is 45-49 awards, whose officers are just below the 50-award threshold to qualify for the Honorable Mention Ribbon award. All estimates control for officer lagged arrests, tenure, unit, and birth year, and include fixed effects for supervisor, year, and month. Wings depict 95% confidence intervals using standard errors clustered by supervisor.

For white officers, we see a discontinuous increase in nomination likelihood for officers in the marginal group—those with 45 to 49 Honorable Mention Certificates. Specifically,

they are 6.5 percentage points ( $p < 0.01$ ) more likely to be nominated for an award relative to officers with 0 to 13 Honorable Mention Certificates. This estimate is statistically significantly different from the estimate for officers with 38-44 Honorable Mention Certificates, lending further support to the theory that white supervisors are advocating for their white officers.

The pattern for Black officers, however, is different. None of the estimates are statistically significant, suggesting that the nomination likelihood does not differ by the number of past awards for Black officers. The Black-white difference is not statistically significant in any of the categories except the last category (officers with 50 or more Honorable Mention Certificates). These findings lend support to our theory that the Black-white nomination gap may be a result of a Black-white advocacy gap.

## 6 Experimental Evidence

The previous section used CPD administrative data to present evidence suggesting that a racial attention gap may lead to a racial nomination gap. In this section, we explore this mechanism in an online experiment that seeks to examine whether evaluators choose to learn about minority officers.<sup>49</sup>

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.<sup>50</sup> We ran an online experiment to isolate the psychic cost of engaging with people of different races. Specifically, we examine whether participants are

---

<sup>49</sup>The experiment was pre-registered in the AEA RCT Registry, AEARCTR-0005929.

<sup>50</sup>For supervisors who do not exhibit racial bias, the psychic cost would be orthogonal to officer race.

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.<sup>51</sup> Officer profiles displayed only demographic information, and participants had to 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 measure of evaluator engagement.

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.<sup>52</sup>

## 6.1 Experimental Design

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

In the first type of task, participants chose between a Black officer and a non-Black officer, where the Black officer was randomly assigned to be either “high-quality” or “low-quality” and the non-Black officer was assigned the converse.<sup>53</sup> “High-quality” profiles

---

<sup>51</sup>This task was incentive compatible as participants were informed that their nomination decisions would be shared with CPD. After the experiment concluded, we 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.

<sup>52</sup>In an experiment, Dickinson et al. (2015) finds that police commissioners are slightly more likely than non-police subjects to issue rewards but with less intensity. However, these differences are not statistically significant.

<sup>53</sup>There were two versions of this task. In the first version, officer race and officer sex were independently chosen. In the second, we fixed the sex to be the same across the two profiles (e.g., either both profiles were male or both were female).

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.<sup>54</sup> In this task, all of the officer’s demographic and work performance measures were visible to the evaluator. See Appendix Figure B1 for an example screenshot.

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.<sup>55</sup> All officers were of “average quality”, defined as having zero or one civilian complaints and an average number of arrests. There were two iterations of this task. In the first iteration, three of the four profiles were always white officers and the race of the fourth profile was randomly chosen amongst white, Black, and Hispanic. In the second iteration, the officer group was racially heterogeneous. Three of the four profiles always featured a white officer, a Black officer, and an Hispanic officer. The race of the fourth profile was randomly chosen amongst these three races. The ordering of these two iterations was randomized.

The ordering of the two 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).<sup>56</sup> For the second type of task, participants were restricted from uncovering any work performance measures for ten seconds and they were also prevented from nominating an officer during these ten seconds. This was to ensure that participants

---

<sup>54</sup>By coupling the number of arrests with the number of civilian complaints, we hoped to signal a more accurate measure of policing quality as opposed to using just one performance measure. Although this is still a simplified proxy for policing quality, we wanted to be cognizant of the fact that we were requiring MTurk evaluators to make decisions in a limited amount of time and did not want to overload them with too much information.

<sup>55</sup>See Appendix Figure B2 for an example screenshot.

<sup>56</sup>These time limits appear to be within reason; participants took about 9.8 seconds, on average, for the pairwise comparison and 27.9 seconds, on average, for the group comparison. For the group comparison, conditional on mousing over any profile, about 70 percent of participants moused over all four profiles.



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 doing so. See Appendix B for more information about the online experiment, including balance tests.

## 6.2 Sample Selection and Data

The experiment was conducted on Amazon Mechanical Turk in July 2020 using Qualtrics surveys.<sup>57</sup> We recruited 411 MTurk workers (hereafter “evaluators”) who were 18 years of age or older, based in the United States with English language proficiency, and who had access to a computer with a mouse and Javascript. The technical requirements were necessary in order to capture mouse movements on the screen. The survey had three data quality checks to identify bots and to ensure evaluators paid attention during the survey. For the analysis, we decided to include evaluators who passed at least two of the three data quality checks. This restriction reduces our final analysis sample to 407 evaluators.

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 younger than the average CPD supervisor. The modal MTurk evaluator is in the 26 to 35 age group, while the average CPD supervisor is 44 years old at the start of our analysis sample.

---

<sup>57</sup>It is possible that the George Floyd incident on May 25, 2020 and subsequent protests may have altered people’s perceptions of the police and Black individuals. Specifically, the incident may have increased MTurk workers’ interest in and affinity towards Black officers because they are Black. This would work against our results, which find that Black officers are less likely to be moused over and are less likely to be nominated when paired against a non-Black officer.

### 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 7 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.

Table 7: Impact of being Black on Nomination Likelihood

Pairwise Comparison: Evaluator Race:	Outcome Variable: Nominated Black v. Non-Black			High v. Low
	All	White	Black	All
	(1)	(2)	(3)	(4)
High-Quality Profile	0.467*** (0.0929)	0.393*** (0.0955)	0.763*** (0.290)	0.524*** (0.169)
Black Officer	-0.0848*** (0.0311)	-0.0976*** (0.0362)	-0.0166 (0.0826)	-0.0324 (0.119)
High-Quality x Black Officer	0.0296 (0.0449)	0.0430 (0.0520)	-0.00289 (0.121)	0.141 (0.171)
Female Officer				-0.0705 (0.0680)
High-Quality x Female Officer				0.137 (0.107)
Observations	1,576	1,196	256	794

Source: MTurk survey data.

Notes: This table reports estimates from a pairwise comparison of officer profiles. Columns 1-3 are a pairwise comparison between a Black officer and a non-Black 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 sex, age, tenure, arrests, and complaints. Robust standard errors are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Low-quality Black officers are 8.5 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 about 47 percentage points more likely to be nominated for an award ( $p < 0.01$ ) than low-quality officers, high-quality Black officers are still 5.5 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 -9.8 percentage points ( $p < 0.01$ ) for low-quality officer profiles and -5.5 percentage points for high-quality officer profiles (p-value = 0.111). There is no statistically significant Black-white nomination gap among Black evaluators.

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.

## 6.4 Do evaluators choose to learn about Black officers?

Next, we examine whether MTurk evaluators are differentially less likely to engage with Black profiles, after controlling for the profile location on the computer screen and the evaluator's starting mouse position.

Evaluators tend to mouse over most of the officer profiles: over 80 percent of officer profiles were moused over. 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 of -0.029 is borderline significant, with a p-value of 0.104. The Hispanic-white difference is not statistically significant (p-value: 0.158).

When the officer pool is predominantly white (75% white) the Black-white engagement gap widens to -7.1 percentage points ( $p < 0.1$ ). However, if the minority officer is Hispanic, then there is no statistically significant difference in mouse-over likelihood. We see very similar estimates when we focus on an officer pool where half of the officers are white, an environment that is demographically similar to the Chicago Police Department. When there is a more heterogeneous officer pool (25-50 percent white), the Black-white

difference shrinks to 2.7 percentage points and is not statistically significant. In summary, these results lend support to our theory that a racial attention gap may in part explain the racial advocacy gap.

Table 8: Impact of Officer Race on Evaluator Engagement

Officer Pool:	All	Predom. White	Het. Race, 50% White	Het. Race, 25-50% White
	(1)	(2)	(3)	(4)
Black Officer	-0.0289 (0.0177)	-0.0707* (0.0386)	-0.0807* (0.0433)	-0.0271 (0.0230)
Observations	2,992	1,492	504	1,500
Mean Outcome for Reference Group	0.842	0.838	0.853	0.852

Source: MTurk survey data.

Notes: This table reports estimates for racial differences in evaluator engagement. All estimates control for profile location on screen and evaluator’s starting mouse position. Robust standard errors are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

## 7 Conclusion

This paper studies whether a Black-white gap in award nominations may arise 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 understand the underlying mechanism, we conduct an online experiment wherein we experimentally examine whether evaluators choose to learn about minority officers and find support for the argument that a racial nomination gap may arise in part because supervisors are not choosing to engage with Black officers.

To provide stronger evidence on this mechanism, we examine how the Black-white nomination gap changes when the supervisor is required to learn more information about his officers due to the annual evaluation. White supervisors are more likely to nominate both white and Black officers in the evaluation quarter relative to the baseline quarter, suggesting that the Black-white nomination gap at least partially reflects statistical discrimination. However, a constant, negative Black-white nomination gap across all quarters suggests that white supervisors are not choosing to advocate for white and Black officers equally. For example, the nomination likelihood for white officers increases prior to the evaluation quarter and remains elevated after the evaluation, but neither of this is true for Black officers. Further, we find that white supervisors are discontinuously more likely to nominate white officers who are on the margin of eligibility for an award, but not so for Black officers. Collectively, these results suggest that supervisor bias in advocacy decisions, rather than statistical discrimination, is driving the racial nomination gap.

The implications from our findings are significant because awards help Black officers differentially more than white officers when it comes to career advancement and merit promotions (Appendix Table A1). Yet, Black officers are less likely to be nominated for and win department awards. This racial advocacy gap can have amplifying effects in a setting where supervisors are not randomly assigned. Indeed, we see this when we estimate the Black-white nomination gap among all sergeants as opposed to assigned supervisors. One concern about estimating an unbiased Black-white nomination gap without random supervisor assignment is that both Black and white officers may sort to their preferred supervisors, resulting in a zero nomination gap. But that is not what we see; the Black-white gap more than doubles (Appendix Table A11). Similarly, we find that the Black-white nomination gap is the largest among supervisors with the highest nomination propensity (Appendix Figure A4). That is, supervisors who are the most likely to nominate officers for awards are also the least likely to nominate Black officers relative to white officers.<sup>58</sup>

---

<sup>58</sup>This estimate is statistically significantly different from the other propensity quartiles.

A nascent literature provides empirical evidence that discrimination can be reversed (Ayalew et al., 2021; Bohren et al., 2019; Lowe, 2019; Mousa, 2020).<sup>59</sup> 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 white managers do not, and thus, a racial attention and advocacy gap may lead to 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.

---

<sup>59</sup>The 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.

## 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.
- 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. A., D. Knox, J. Mummolo, and R. Rivera (2021). The role of officer race and gender in police-civilian interactions in chicago. *Science* 371(6530), 696–702.
- 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).
- Cohen, G. L. and J. Garcia (2005). ”I Am Us”: Negative Stereotypes as Collective Threats. *Journal of Personality and Social Psychology* 89(4), 566.

- 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.

- Spencer, S. J., C. Logel, and P. G. Davies (2016). Stereotype threat. *Annual review of psychology* 67, 415–437.
- Srivastava, S. B. and E. L. Sherman (2015). Agents of change or cogs in the machine? re-examining the influence of female managers on the gender wage gap. *American Journal of Sociology* 120(6), 1778–1808.
- U.S. Department of Justice (January 13, 2017). Investigation of the Chicago Police Department. Technical report.
- Van Maanen, J. (1975). Police socialization: A longitudinal examination of job attitudes in an urban police department. *Administrative science quarterly*, 207–228.
- 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.
- 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: Impact of Awards on Merit Promotions

	(1)	(2)
<i>Panel A: Award Receipt</i>		
Awards (in standard deviations)	0.0430** (0.0192)	0.0345** (0.0143)
Black Officer		0.0998** (0.0487)
Awards $\times$ Black Officer		0.116 (0.0921)
Observations	362	362
<i>Panel B: Award Nomination</i>		
Nominations (in standard deviations)	0.0388** (0.0180)	0.0335** (0.0144)
Black Officer		0.0869** (0.0440)
Nominations $\times$ Black Officer		0.0640 (0.0808)
Observations	362	362
Mean Outcome for Reference Group	0.036	0.015

Notes: This table reports estimates for the impact of awards on the likelihood of a merit promotion conditional on passing the Sergeant exam. Panel A reports estimates for award receipt, while Panel B reports estimates for award nominations. The sample is at the officer level. All estimates control for officer race, birth year, tenure at the end of the sample period, and modal unit. Robust standard errors are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A2: Department Awards

	<b>Award Name and Description</b>	<b>Share</b>
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 A3: Racial Difference in Nomination Likelihood by Arrests: All Supervisor Assignments

Specification:	Outcome Variable: Nominated	
	Original Model (1)	All Assignments (2)
Two to three arrests last month	0.0146*** (0.00171)	0.0227*** (0.00225)
Four or more arrests last month	0.0457*** (0.00388)	0.0626*** (0.00478)
<i>Interactions with Black Officer</i>		
Two to three arrests last month	-0.00727*** (0.00247)	-0.0134*** (0.00288)
Four or more arrests last month	-0.0177*** (0.00530)	-0.0224*** (0.00664)
<i>Interactions with Hispanic Officer</i>		
Two to three arrests last month	0.000838 (0.00319)	-0.00254 (0.00291)
Four or more arrests last month	-0.0121** (0.00552)	-0.0133** (0.00536)
Observations	176,552	288,485
Mean Pr(Nom) for Reference Group	0.013	0.017

Notes: This table reports estimates for the impact of an officer's lagged arrest record on the probability of nomination by white supervisors. Column 1 reports the original specification, which was restricted to supervisor-officer assignments that lasted exactly one year. Column 2 expands the sample to all supervisor-officer assignments. 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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

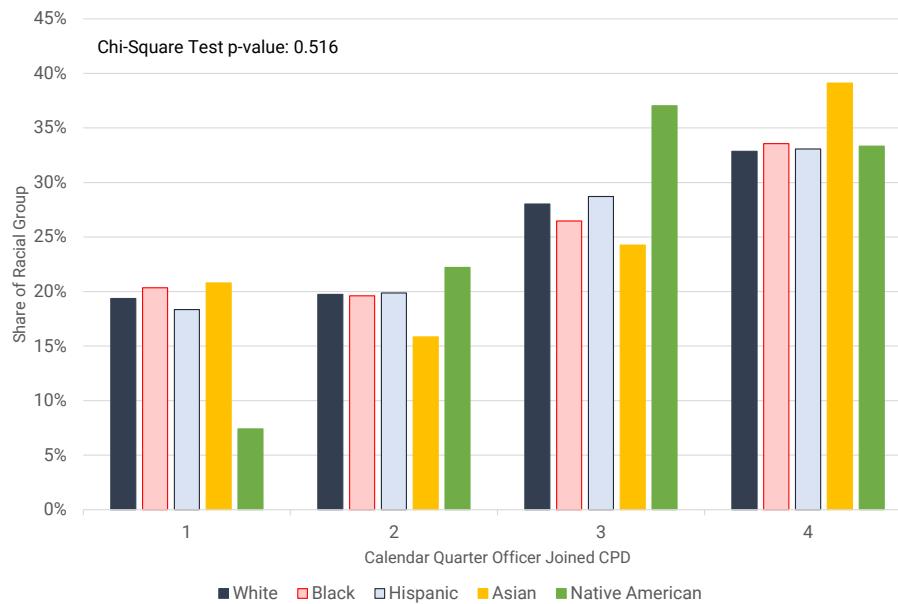
Table A4: Racial Difference in Nomination Likelihood by Quarter: All Supervisor Assignments

Specification:	Outcome Variable: Nominated			
	Original Model	All Assignments		
		All Years	First Year	Subsequent Years
	(1)	(2)	(3)	(4)
Black Officer	0.00358 (0.00363)	-0.00332 (0.00405)	0.0226** (0.0106)	-0.0259*** (0.00953)
<i>Quarter relative to two quarters before evaluation</i>				
Two quarters pre-evaluation	0.0116*** (0.00308)	0.00792** (0.00317)	0.0251*** (0.00876)	0.000127 (0.00693)
One quarter pre-evaluation	0.0175*** (0.00322)	0.0118*** (0.00338)	0.0341*** (0.00930)	-0.00393 (0.00786)
Evaluation quarter	0.0207*** (0.00368)	0.0131*** (0.00347)	0.0464*** (0.0109)	-0.0143* (0.00772)
One quarter post-evaluation	0.0133*** (0.00383)	0.00875** (0.00388)	0.0427*** (0.0122)	-0.00969 (0.00905)
Two quarters post-evaluation	0.0111*** (0.00415)	0.00411 (0.00424)	0.0452*** (0.0121)	-0.0265*** (0.00997)
Three quarters post-evaluation	0.00689 (0.00481)	0.000462 (0.00510)	0.0394*** (0.0129)	-0.0261** (0.0127)
<i>Two quarters pre-evaluation x</i>				
Black Officer	-0.00971** (0.00383)	-0.00652* (0.00394)	-0.0293** (0.0114)	0.0146 (0.00945)
Hispanic Officer	-0.00574 (0.00486)	-0.00279 (0.00539)	-0.00536 (0.0126)	-0.00160 (0.0142)
<i>One quarter pre-evaluation x</i>				
Black Officer	-0.0130*** (0.00372)	-0.00719* (0.00403)	-0.0194 (0.0118)	0.00966 (0.00994)
Hispanic Officer	-0.00479 (0.00512)	-2.82e-05 (0.00532)	0.00651 (0.0130)	-0.00343 (0.0140)
<i>Evaluation quarter x</i>				
Black Officer	-0.0136*** (0.00399)	-0.00790* (0.00422)	-0.0360*** (0.0125)	0.0163 (0.0106)
Hispanic Officer	-0.00480 (0.00543)	-0.00158 (0.00561)	-0.0103 (0.0131)	-0.00105 (0.0147)
<i>One quarter post-evaluation x</i>				
Black Officer	-0.0139*** (0.00392)	-0.00734* (0.00427)	-0.0241** (0.0122)	0.0144 (0.0118)
Hispanic Officer	-0.00864* (0.00500)	-0.00491 (0.00550)	-0.00206 (0.0122)	-0.00847 (0.0152)
<i>Two quarters post-evaluation x</i>				
Black Officer	-0.0130*** (0.00410)	-0.00314 (0.00476)	-0.0275* (0.0143)	0.0290** (0.0122)
Hispanic Officer	-0.00775 (0.00533)	-0.00115 (0.00599)	0.00978 (0.0147)	-0.00606 (0.0162)
<i>Three quarters post-evaluation x</i>				
Black Officer	-0.0123** (0.00521)	-0.00338 (0.00581)	-0.0294* (0.0153)	0.0261* (0.0158)
Hispanic Officer	-0.0140** (0.00591)	-0.00891 (0.00687)	-0.0157 (0.0163)	-0.000824 (0.0188)
Observations	176,552	288,485	41,862	64,896
Mean Pr(Nom) for Reference Group	0.019	0.046	0.041	0.104

Notes: The table depicts how the probability of nomination changes by quarter relative to three quarters before the officer's evaluation. Column 1 reports the original specification, which was restricted to supervisor-officer assignments that lasted exactly one year. Columns 2-4 expand the sample to all supervisor-officer assignments. 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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$



Figure A1: Evaluation Quarter by Officer Race



Source: CPD analysis sample.

Notes: This figure displays the share of officers in each CPD start quarter, which determines the evaluation quarter, by officer race. The p-value of a chi-square test of independence across the race categories is 0.516.

Table A5: 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 A6: Supervisor Race and Lagged Officer Work Measures: Analysis Sample

Outcome Variable: Supervisor is White		
Coefficients for:	White Officer (1)	Black-White Difference (2)
Violent-crime Arrests	0.00147 (0.00142) [1.000]	0.00298 (0.00200) [1.000]
Property-crime Arrests	0.000659 (0.00133) [1.000]	3.87e-05 (0.00170) [1.000]
Non Index-crime Arrests	0.000683** (0.000298) [0.593]	-0.000697 (0.000527) [1.000]
Complaints	1.90e-05 (0.00431) [1.000]	0.00858 (0.0102) [1.000]
TRR Filings	-0.0114 (0.00768) [1.000]	-0.000401 (0.0149) [1.000]
Strong Force Ratio	0.000391 (0.0146) [1.000]	-0.00628 (0.0300) [1.000]
Observations		19,423
R-squared		0.075
p-value for joint F-test	0.128	0.434

Notes: The table reports estimates for lagged annual work measures on the likelihood of being assigned to a white supervisor. The sample is restricted to all supervisor-officer relationships that last one year (“analysis sample”). 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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A7: Nomination Likelihood by Lagged Arrests and Officer Race: Additional Controls

	Outcome Variable: Nominated				
	(1)	(2)	(3)	(4)	(5)
Two to three arrests last month	0.0146*** (0.00171)	0.0145*** (0.00172)	0.0143*** (0.00173)	0.0143*** (0.00173)	
Four or more arrests last month	0.0457*** (0.00388)	0.0453*** (0.00388)	0.0449*** (0.00390)	0.0449*** (0.00390)	
Index-crime arrests: 2-3 arrests last month					0.00565** (0.00258)
Index-crime arrests: 4 or more arrests last month					0.0170*** (0.00596)
Non-Index crime arrests: 2-3 arrests last month					0.0184*** (0.00260)
Non-Index crime arrests: 4 or more arrests last month					0.0547*** (0.00589)
<i>Interactions with Black Officer</i>					
Two to three arrests last month	-0.00727*** (0.00247)	-0.00732*** (0.00247)	-0.00730*** (0.00247)	-0.00728*** (0.00247)	
Four or more arrests last month	-0.0177*** (0.00530)	-0.0177*** (0.00531)	-0.0177*** (0.00531)	-0.0177*** (0.00531)	
Index-crime arrests: 2-3 arrests last month					-0.00198 (0.00354)
Index-crime arrests: 4 or more arrests last month					-0.0189*** (0.00723)
Non-Index crime arrests: 2-3 arrests last month					-0.00764* (0.00432)
Non-Index crime arrests: 4 or more arrests last month					-0.0145 (0.00901)
<i>Interactions with Hispanic Officer</i>					
Two to three arrests last month	0.000838 (0.00319)	0.000859 (0.00319)	0.000866 (0.00319)	0.000850 (0.00319)	
Four or more arrests last month	-0.0121** (0.00552)	-0.0121** (0.00551)	-0.0121** (0.00551)	-0.0122** (0.00551)	
Index-crime arrests: 2-3 arrests last month					-0.000937 (0.00381)
Index-crime arrests: 4 or more arrests last month					-0.00400 (0.00873)
Non-Index crime arrests: 2-3 arrests last month					0.00176 (0.00449)
Non-Index crime arrests: 4 or more arrests last month					-0.0130 (0.00839)
Observations	176,552	176,552	176,552	176,552	176,552
Mean Pr(Nom) for Reference Group	0.013	0.013	0.013	0.013	0.014
Controls for:					
Complaints		Yes	Yes	Yes	Yes
TRR Filings			Yes	Yes	Yes
Strong Force Ratio				Yes	Yes

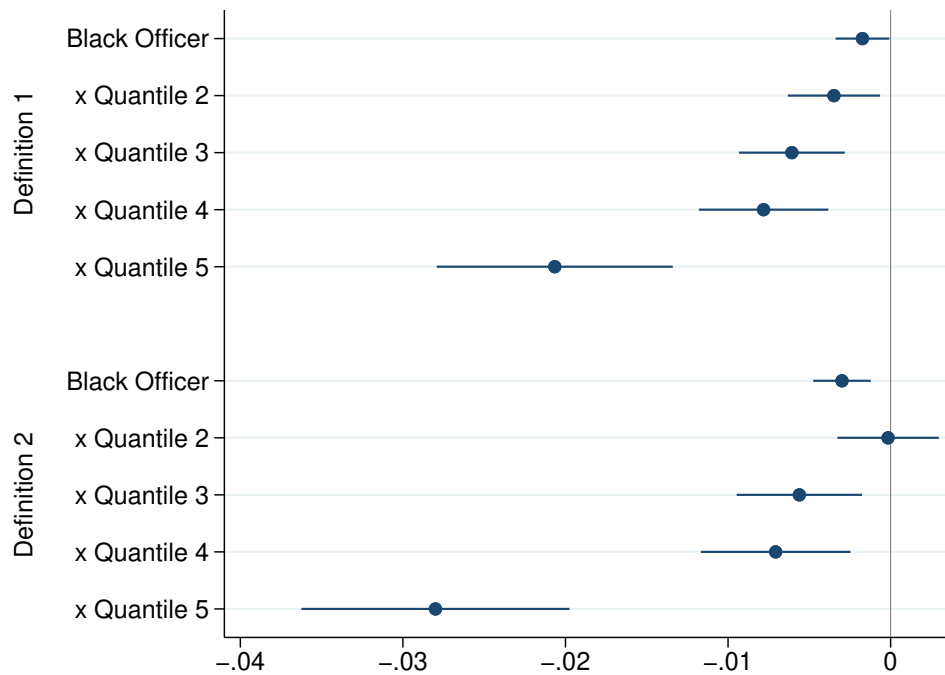
Notes: This table reports estimates for the impact of an officer's lagged arrest record on the probability of nomination by white supervisors. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year and tenure. Standard errors clustered by supervisor are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A8: Nomination Likelihood by Lagged Arrests and Officer Race: Robustness Checks

Specification: Standard errors clustered by:	Outcome Variable: Nominated				
	Original Model Supervisor (1)	Officer FE Supervisor (2)	Sup + Off FE Supervisor (3)	Officer FE Officer (4)	Sup + Off FE Officer (5)
Two arrests last month	0.0146*** (0.00171)	0.00785*** (0.00173)	0.00743*** (0.00174)	0.00785*** (0.00174)	0.00743*** (0.00173)
Three or more arrests last month	0.0457*** (0.00388)	0.0238*** (0.00284)	0.0202*** (0.00265)	0.0238*** (0.00275)	0.0202*** (0.00251)
<i>Interactions with Black Officer</i>					
Two to three arrests last month	-0.00727*** (0.00247)	-0.00400 (0.00244)	-0.00381 (0.00244)	-0.00400 (0.00243)	-0.00381 (0.00245)
Four or more arrests last month	-0.0177*** (0.00530)	-0.00960** (0.00470)	-0.00727 (0.00457)	-0.00960** (0.00478)	-0.00727 (0.00449)
<i>Interactions with Hispanic Officer</i>					
Two to three arrests last month	0.000838 (0.00319)	0.00253 (0.00303)	0.00321 (0.00300)	0.00253 (0.00301)	0.00321 (0.00297)
Four or more arrests last month	-0.0121** (0.00552)	-0.00532 (0.00429)	-0.00573 (0.00415)	-0.00532 (0.00499)	-0.00573 (0.00452)
Observations	176,552	176,552	176,552	176,552	176,552
Mean Pr(Nom) for Reference Group	0.013	0.013	0.013	0.013	0.013
Controls for:					
Supervisor	Yes		Yes	Yes	Yes
Officer		Yes	Yes		Yes

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. Clustered standard errors are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Figure A2: Black-White Nomination Gap by Predicted Nomination Probability



Source: CPD analysis sample.

Notes: This figure reports the Black-white nomination gap by quantile of predicted nomination probability. The predicted nomination probability was obtained from an out-of-sample regression of a binary nominated outcome variable on lagged violent-crime arrests, lagged property-crime arrests, lagged non-index-crime arrests, lagged TRR filings, lagged complaints, lagged ratio of strong use of force, birth year, tenure, unit, and month. "Defintion 1" uses 2009, 2010, and 2011 as out-of-sample years. "Defintion 2" uses 2009, 2011, and 2013 as out-of-sample years.

Table A9: Impact of Lagged Arrests on Nomination Likelihood by Supervisor Race

Supervisor Race:	Outcome Variable: Nominated		
	White (1)	Black (2)	Hispanic (3)
Two to three arrests last month	0.0146*** (0.00171)	0.00720* (0.00372)	0.0109*** (0.00325)
Four or more arrests last month	0.0457*** (0.00388)	0.0336*** (0.00811)	0.0407*** (0.00725)
<i>Interactions with Black Officer</i>			
Two to three arrests last month	-0.00727*** (0.00247)	-0.00337 (0.00472)	0.00244 (0.00680)
Four or more arrests last month	-0.0177*** (0.00530)	-0.0100 (0.0136)	-0.0122 (0.0143)
<i>Interactions with Hispanic Officer</i>			
Two to three arrests last month	0.000838 (0.00319)	0.00156 (0.00937)	0.000892 (0.00483)
Four or more arrests last month	-0.0121** (0.00552)	-0.0183* (0.0105)	-0.00608 (0.0108)
Observations	176,552	30,476	37,436
Mean Pr(Nom) for Reference Group	0.013	0.006	0.011

Notes: This table reports estimates for the impact of an officer's lagged arrest record on the probability of nomination by white supervisors (column 1), Black supervisors (column 2), and Hispanic supervisors (column 3). 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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

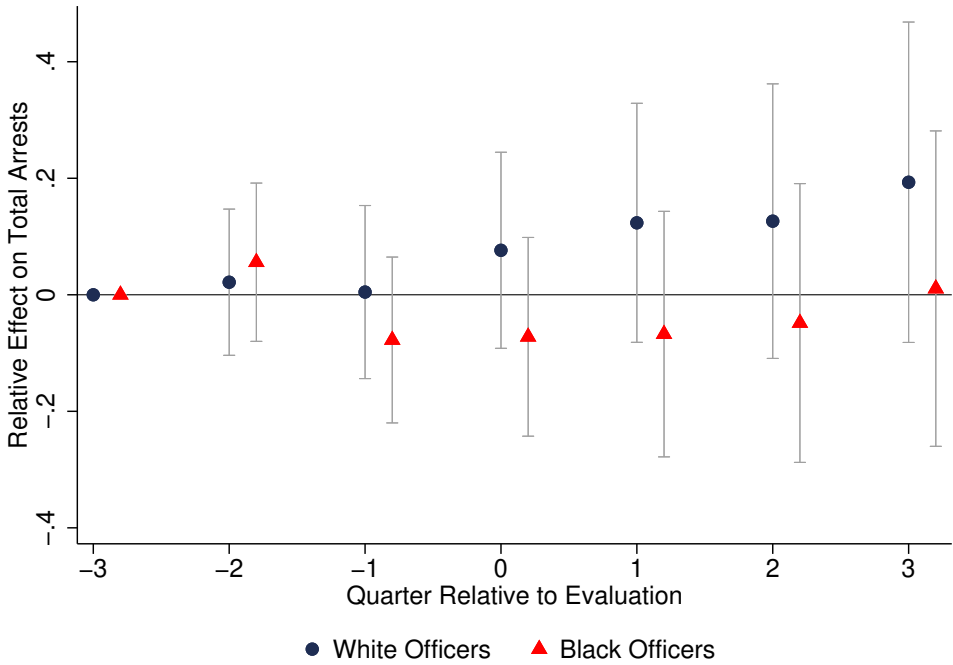
Table A10: Racial Difference in Nomination Likelihood by Quarter and Supervisor Race

Supervisor Race:	Outcome Variable: Nominated		
	White (1)	Black (2)	Hispanic (3)
Black Officer	0.00358 (0.00363)	0.00313 (0.00734)	-0.00137 (0.00639)
<i>Quarter relative to two quarters before evaluation</i>			
Two quarters pre-evaluation	0.0116*** (0.00308)	0.0106 (0.00847)	0.00933 (0.00644)
One quarter pre-evaluation	0.0175*** (0.00322)	0.0134* (0.00766)	0.0219*** (0.00748)
Evaluation quarter	0.0207*** (0.00368)	0.0133* (0.00800)	0.0212*** (0.00695)
One quarter post-evaluation	0.0133*** (0.00383)	-0.000591 (0.00830)	0.0118* (0.00688)
Two quarters post-evaluation	0.0111*** (0.00415)	0.00582 (0.0114)	0.0113 (0.00820)
Three quarters post-evaluation	0.00689 (0.00481)	-0.00285 (0.00931)	0.00183 (0.00839)
<i>Two quarters pre-evaluation ×</i>			
Black Officer	-0.00971** (0.00383)	-0.0115 (0.00860)	-0.00416 (0.00840)
Hispanic Officer	-0.00574 (0.00486)	-0.0125 (0.0101)	0.00105 (0.0104)
<i>One quarter pre-evaluation ×</i>			
Black Officer	-0.0130*** (0.00372)	-0.0112 (0.00734)	-0.00915 (0.00766)
Hispanic Officer	-0.00479 (0.00512)	0.00153 (0.0108)	-0.00828 (0.0102)
<i>Evaluation quarter ×</i>			
Black Officer	-0.0136*** (0.00399)	-0.0116 (0.00803)	-0.00137 (0.00801)
Hispanic Officer	-0.00480 (0.00543)	0.00925 (0.0121)	0.00210 (0.0104)
<i>One quarter post-evaluation ×</i>			
Black Officer	-0.0139*** (0.00392)	-0.00121 (0.00785)	0.00214 (0.00772)
Hispanic Officer	-0.00864* (0.00500)	0.00924 (0.0105)	-0.000356 (0.00858)
<i>Two quarters post-evaluation ×</i>			
Black Officer	-0.0130*** (0.00410)	-0.0107 (0.0103)	0.00916 (0.00999)
Hispanic Officer	-0.00775 (0.00533)	0.00650 (0.0104)	-0.0108 (0.00986)
<i>Three quarters post-evaluation ×</i>			
Black Officer	-0.0123** (0.00521)	-0.00666 (0.00814)	0.0128 (0.00882)
Hispanic Officer	-0.0140** (0.00591)	-0.000624 (0.0143)	-0.00933 (0.0118)
Observations	176,552	30,476	37,436
Mean Pr(Nom) for Reference Group	0.019	0.018	0.019

Notes: The table depicts how the quarterly probability of nomination changes relative to three quarters before the officer's evaluation for white supervisors (column 1), Black supervisors (column 2), and Hispanic supervisors (column 3). 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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$



Figure A3: Number of Arrests by Relative Quarter and Officer Race



Source: CPD analysis sample.

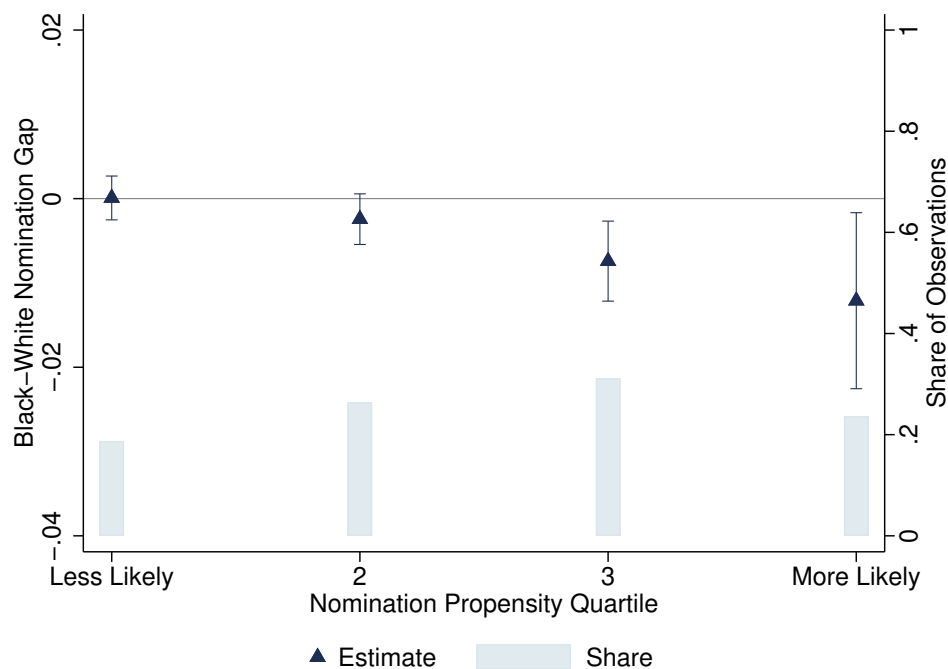
Notes: This figure depicts how the quarterly number of arrests changes relative to three quarters before the officer's evaluation, separately for white and Black officers. All estimates include supervisor, unit, month, and year fixed effects, and control for officer birth year and tenure. Wings depict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

Table A11: Impact of Lagged Arrests on Nomination Likelihood by Nominating Source

Nominations by:	Outcome Variable: Nominated		
	Current Supervisor (1)	Everyone (2)	Non-Sup. Sergeants (3)
Two to three arrests last month	0.0146*** (0.00171)	0.0885*** (0.00394)	0.0645*** (0.00347)
Four or more arrests last month	0.0457*** (0.00388)	0.165*** (0.00598)	0.110*** (0.00517)
<i>Interactions with Black Officer</i>			
Two to three arrests last month	-0.00727*** (0.00247)	-0.0393*** (0.00581)	-0.0247*** (0.00532)
Four or more arrests last month	-0.0177*** (0.00530)	-0.0438*** (0.00907)	-0.0315*** (0.00757)
<i>Interactions with Hispanic Officer</i>			
Two to three arrests last month	0.000838 (0.00319)	-0.0101* (0.00579)	-0.00782 (0.00502)
Four or more arrests last month	-0.0121** (0.00552)	-0.0130* (0.00683)	-0.00479 (0.00590)
Observations	176,552	288,485	288,485
Mean Pr(Nom) for Reference Group	0.013	0.126	0.099

Notes: This table reports estimates for the impact of an officer's lagged arrest record on the probability of nomination by white supervisors. Column 1 reports the original specification from Table 5 column 1, nominations by the officer's current supervisor. Column 2 reports nominations by all nominating sources. Column 3 reports nominations by any sergeant who is neither a current supervisor nor a past supervisor. 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. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Figure A4: Black-White Nomination Gap by Nomination Propensity



Source: CPD analysis sample.

Notes: This figure depicts how the Black-white nomination gap changes by nomination propensity. The nomination propensity 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. 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.

## 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 B3 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.

Figure B1: Screenshot of Pairwise Comparison Task

06

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

☐

Initials	A.L.
Race	White
Sex	Male
Age	51
Experience	9.33
Total arrests	24
Civilian complaints	1

☐

Initials	R.N.
Race	Black
Sex	Male
Age	47
Experience	8.08
Total arrests	35
Civilian complaints	0

Figure B2: 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.

☐

Initials	K.B.
Race	White
Sex	Male
Age	28
Experience	
Total arrests	
Civilian complaints	

☐

Initials	D.S.
Race	Black
Sex	Male
Age	38
Experience	
Total arrests	
Civilian complaints	

☐

Initials	S.D.
Race	Hispanic
Sex	Male
Age	32
Experience	
Total arrests	
Civilian complaints	

☐

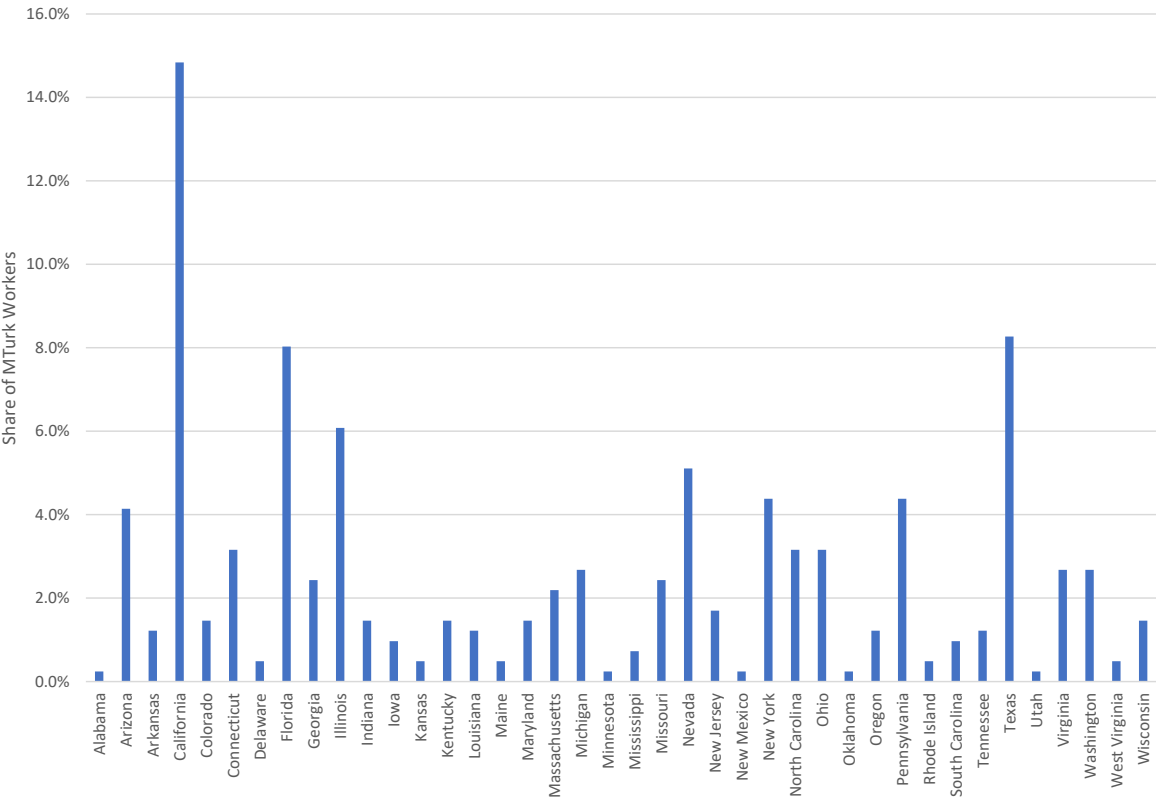
Initials	S.O.
Race	White
Sex	Male
Age	41
Experience	
Total arrests	
Civilian complaints	

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 B3: Distribution of MTurk Worker State of Residence



Source: MTurk survey data.

Table B2: Balance Tests for the Pairwise Comparison Task

Version:	Diff Race, Diff Sex		Diff Race, Same Sex	
Officer Profile Quality:	Low (1)	High (2)	Low (3)	High (4)
Age	0.379	0.710	0.367	0.771
Tenure	0.269	0.894	0.523	0.423
Arrests	0.821	0.554	0.317	0.945

Source: CPD profiles for MTurk online experiment.

Notes: This table reports p-values from a t-test or regression of means between Black profiles and non-Black profiles on officer traits. Officer complaints were not included because there is no variation within the high/low categories. For example, high-quality profiles have zero complaints by construction and low-quality profiles have one or two complaints by construction.

Table B3: Balance Tests for the Group Comparison Task

Officer Pool:	Predom. White (1)	Het. Race (2)
Age	0.581	0.362
Tenure	0.959	0.282
Arrests	0.265	0.738
Complaints	0.313	0.633

Source: CPD profiles for MTurk online experiment.

Notes: This table reports p-values from a t-test of means between white profiles and non-white profiles (column 1) and Black profiles and non-Black profiles (column 2).