In-group Bias and the Police: Evidence from Chicago[*]

Nayoung Rim[†]   Bocar Ba[‡]   Roman Rivera[§]

July 9, 2019

Abstract

This paper examines the impact of in-group bias on the internal dynamics of a police department. Prior studies have documented racial bias in policing, but little is known about bias against officers due to lack of available data. We construct a novel panel dataset of Chicago Police Department officers, with detailed information on officer characteristics and work productivity. Exploiting quasi-random variation in supervisor assignment, we find that white supervisors are less likely to nominate black officers than white or Hispanic officers. We find weaker evidence that male supervisors are less likely to nominate female officers than male officers. We explore several theories of discrimination that can explain our main findings. The evidence is most consistent with a theory of animus towards black officers.

JEL Codes: J71, J15, J48
Keywords: bias, police officers, award nominations, supervisors

[†]United States Naval Academy, Department of Economics. Email: `rim@usna.edu`
[‡]University of Pennsylvania Law School. Email: `bocarba@law.upenn.edu`
[§]Columbia University, Department of Economics. Email: `rgr2122@columbia.edu`

1   Introduction

The recent high-profile cases of unarmed black men fatally shot by police officers raise the question of whether racial bias exists in police use of force (Nix et al. 2017; Fryer 2018; Knox, Lowe, and Mummolo 2019).[1] Even in more commonplace interactions, such as receiving speeding tickets and citations for traffic violations, a growing literature is uncovering racial bias against minority civilians (West 2018; Goncalves and Mello 2018; Coviello and Persico 2015). These racial disparities in law enforcement present an important social policy concern in the United States, particularly as there need not be a trade-off between effectiveness and fairness in policing (Persico 2002).

One proposed solution is increased diversity in police ranks, which traditionally are overwhelmingly white and male. For example, in their investigative report of the Ferguson Police Department, the U.S. Department of Justice called for a more diverse police force as part of a broader reform effort (United States Department of Justice 2015, p. 58). Several cities, including Chicago, Indianapolis, and Knoxville, have followed this lead and pursued diversity initiatives.[2] In addition to improving policing quality (Donohue III and Levitt 2001; McCrary 2007; Miller and Segal 2018; Hoekstra and Sloan 2019), minority representation, particularly at higher ranks of office, may help to recruit more minorities and close promotion gaps, which may further attract minority applicants. Prior research has found that minorities in management positions can address wage gaps and occupational gaps (Langan 2018; Kofoed and mcGovney 2019).

At the same time, these policies may be constrained by the extent to which officer bias carries over to their colleagues. An extensive literature documents racial and gender bias in the workplace, which may hinder minorities' career progression. For example, Egan, Matvos,

---

[1]See http://time.com/4404987/police-violence/.

[2]Chicago: https://www.chicagotribune.com/news/breaking/ct-met-chicago-police-hiring-20180503-story.html; Indianapolis: https://www.indystar.com/story/opinion/columnists/suzette-hackney/2018/09/27/impd-leads-charge-toward-diversity-columnist-suzette-hackney-writes/1433649002/; Knoxville: https://www.knoxnews.com/story/news/local/2017/01/09/knoxville-police-department-recruits-remain-diverse-group/96345092/

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

The theoretical predictions for law enforcement are unclear. On one hand, studies detecting racial bias in policing imply that officers may engage in similar behavior towards their minority colleagues. On the other hand, the organizational culture of policing is that of a homogenous identity with a uniform goal of reducing crime (Sklansky 2011; Sklansky 2005). This insider identity, that "blue is blue" and the resulting irrelevance of the individual identity of the person in uniform, is shaped by the nature of the job and may limit bias to interactions with "outsider" subjects. Empirically, little is known about in-group bias against officers due to the lack of available data. The primary purpose of this paper is to examine the impact of bias on the internal dynamics of a police department.

We construct a novel panel dataset of all Chicago Police Department (CPD) officers between 2009 and 2015 containing information on officer demographics, employment, and work productivity. This was a huge undertaking involving more than ten separate datasets gathered from different sources. Most importantly, this dataset contains information on officer use of force, arrests, and misconduct—crucial information in an empirical study of bias.

Using supervisor nominations for departmental awards, we test whether police officers are biased against their minority colleagues. Our identification strategy exploits an institutional feature of the Chicago Police Department that assigns a new supervisor every January. Similar to Chetty, Friedman, and Rockoff 2014, which uses quasi-random variation induced by teacher turnover, we use a new supervisor assignment to approximate random assignment

of a supervisor's race or sex to an officer. We confirm this by analyzing supervisor-officer assignments and find that, conditional on their race and sex, minority officers do not sort to majority supervisors based on their productivity measures, suggesting that we should not see any racial or gender differences in nominations in the absence of supervisor bias.

We find that after switching to a white supervisor, white officers are 4.2 percentage-points more likely to be nominated compared to when they were assigned to a non-white supervisor. This equates to a 24% increase and is statistically significant at the 5% level. The negative black-white nomination gap, on the other hand, widens by 3.5 percentage-points, which is nearly 20% of a white officer's baseline nomination probability. We find similar patterns when we define in-group by gender. Male officers are 5 percentage-points or 28% more likely to be nominated after they switch to a male supervisor, and this estimate is significant at the 1% level. The female-male difference, in contrast, is about half this, at -2.4 percentage-points, but is not statistically significant. These results control for officer characteristics, including arrest record and misconduct, and time-varying neighborhood factors, like unit-level crime rates.

Our second analysis exploits the richness of our panel data and examines the impact of officers' productivity on their likelihood of being nominated for an award. We find that there are increasing returns to having more arrests, with a marked increase in nomination likelihood for those with 5 or more arrests. Black officers under white supervisors also see increasing returns to having more arrests, but the return is less. For example, between a black officer and a white officer, each with one arrest, the same years of experience, and in the same unit, the average white supervisor is 79.3% less likely to nominate the black officer for an award. This gap diminishes as the number of arrests increases, but does not go away. For example, the average black officer with 5 or more monthly arrests—who is at the 94th percentile of their distribution—is still 47% less likely to be nominated by a white supervisor compared to the average white officer with 5 or more monthly arrests, who is at the 81st percentile of their distribution.

We explore several mechanisms that can explain our main findings. About half of the minority nominations gap can be explained by officer productivity, neighborhood characteristics, and supervisor interaction, with the other 50% remaining unexplained. The evidence is most consistent with a theory of animus towards black officers. Although it is difficult to prove implicit bias without a scientific measure like the Implicit Association Test, we do find that white supervisors who have a more negative black-white nomination gap—that is, those who are less likely to nominate black officers—are also more likely to use force against black subjects relative to white subjects. Further, this theory is supported by our finding that a negative black-white nomination gap exists even among officers who are at the right-tail of the productivity distribution.

An alternative explanation concerns supervisor-officer interactions. White supervisors may be less likely to nominate black officers because they are less likely to interact with them, and, therefore, are less likely to be informed of their accomplishments. Although we find that lack of supervisor interaction depresses nominations, we do not find differential effects by race.

Last, we explore a channel through which negative stereotypes may indirectly affect the black-white nomination gap. Negative stereotypes may lead to collective threat, or the concern by a minority group member of being devalued by an in-group member's behavior. In this case, black supervisors may perseverate black-white disparities rather than reduce them. The signs and magnitudes are consistent with a theory that black supervisors hold black officers to a higher standard than white officers and white supervisors, though none of the estimates are statistically significant.

Our findings are consistent with studies that find that minorities are less likely to be acknowledged for their work (Hengel 2019; Sarsons 2019) and a literature that establishes the existence of bias among managers and work colleagues (Bertrand and Mullainathan 2004; Giuliano, Levine, and Leonard 2009; Glover, Pallais, and Pariente 2017; Egan, Matvos, and Seru 2018; Sarsons 2018). The evidence is mixed on whether bias is due to animus (taste-

4

based discrimination) or statistical discrimination, as results vary depending on the study setting.[3] Our paper analyzes bias among officers, linking this literature to studies on racial disparities in law enforcement. Like Sarsons 2019, we find that statistical discrimination cannot fully explain supervisor behavior. Consistent with Glover, Pallais, and Pariente 2017, we find that interactions with minority officers are an important mechanism to reduce disparities. But, in contrast with these studies, we find that animus may also be a contributor to the observed bias.

With respect to law enforcement, our study adds to the growing research that finds racial bias in policing. Prior studies largely use data on officer-initiated encounters (Knowles, Persico, and Todd 2001; Anwar and Fang 2006; Antonovics and Knight 2009; Nix et al. 2017; Goncalves and Mello 2018), which may be biased because they do not include the universe of all possible police interactions (Knox, Lowe, and Mummolo 2019). Two notable exceptions are Weisburst 2018, which uses data on officers randomly dispatched to respond to 911 calls, and West 2018, which uses data on officers randomly dispatched to investigate automobile crashes. Similarly, our paper bypasses the truncated data problem by focusing on nomination rates among quasi-randomly assigned officers.

We begin the rest of the paper with a description of our data collection efforts and summary statistics on our analysis sample (Section 2). Section 3 presents our empirical strategy and tests the identifying assumption that supervisors are randomly assigned to officers conditional on race and sex. We present our main results in Section 4, showing that across different analyses, white supervisors are less likely to nominate black officers than white or Hispanic officers. Section 5 discusses several mechanisms that can explain our findings. Section 6 concludes.

---

[3]Additionally, Bohren et al. 2019 suggest the accuracy of the individual's beliefs is important and argues for a distinction between accurate statistical discrimination and inaccurate statistical discrimination.

2   Data

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

2.1   Police Officers 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 unique characteristics available, like 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 the aggregated data, using the most common observation across datasets.[4] Officer rank is taken from salary data provided by the Chicago Department of Human Resources (DHR), covering 2002 to 2017.

Supervisors.   The supervisor dataset provides information about the supervisor who conducted each officer's annual evaluations between 2009 and 2017. Supervisors must be at a higher rank than the officer. Our analysis examines those at the rank of Police Officer, meaning their supervisors are at the rank of Sergeant. In this paper, the term "supervisor" refers to a Sergeant who is officially assigned to conduct a Police Officer's annual evaluation for that calendar year.

Awards.   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 for officer evaluations and merit

---

[4]Not all demographic information is complete in each file, so an aggregation of demographic variables across multiple files is necessary.

promotions.[5] 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).[6]

We consider all performance awards that are open to all sworn Department members.[7] Since we are interested in examining bias towards work colleagues, we further restrict our list to awards that require a superior officer's nomination. To clarify, the superior officer does not need to be an officer's officially assigned supervisor in order to request a nomination. Most awards require a nomination process. An example of an exception is the Attendance Recognition Award, which is determined by an automated process. After these restrictions, our analysis considers 18 awards. Appendix Table A1 provides a description of these awards.

Unit Assignment. The Chicago Police Department is composed of five bureaus: Bureau of Patrol, Bureau of Detectives, Bureau of Organized Crime, Bureau of Organizational Development, and Bureau of Technical Services. These bureaus are further sub-divided into units. 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.

The Bureau of Patrol is responsible for "general field operations, including the protection of life and property; apprehension of criminals; and enforcement of traffic laws and ordinances (Chicago Police Department, Department Organization Directive, General Order G01-02-03)."[8] This bureau is composed of Bureau of Patrol Areas, which are 25 geographic police units spanning the city of Chicago[9], and a Special Functions Division, whose mission is to

---

[5]Chicago Police Department, Career Development Directive, Employee Resource E05-01, retrieved from http://directives.chicagopolice.org/directives/data/a7a56e3d-12887ea9-ce512-887e-c3dce7cd73e28d57.html?ownapi=1

[6]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.

[7]Most awards fit this category (21 out of 35). 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.

[8]Retrieved from http://directives.chicagopolice.org/directives/data/a7a57be2-1291da66-88512-91e6-521a90347177e975.html

[9]Between 2012-2014, units 13, 21, and 23 were dissolved and merged with neighboring units. See Appendix

support Bureau of Patrol operations. Examples of Special Functions Division within the Bureau of Patrol are the special weapon and tactics (SWAT) unit, mounted unit, and canine unit. Upon graduation from Police Academy, all Police Officers begin their career in one of the 25 geographic units.

Arrests. The arrests dataset contains information on all arrests made by Department members. The dataset includes detailed information about the subject, crime, and arrest location and time. These data cover 2001-2017 but arrest day and month are only provided from 2010 onwards. For arrests made in 2009, we use the date the subject was released to Cook County jail as a proxy for the arrest date.[10] For our analysis, we use total officer arrests as well as arrests in three aggregated crime categoires: 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. Violent crimes are crimes related to violence, such as murder and assault. Property crimes are crimes related to property, such as burglary and motor vehicle theft. Non-Index crimes, on the other hand, capture all other categories, such as municipal code violations, traffic violations, warrants, drugs, prostitution, gambling, etc.[11]

Complaints. The complaints data contain all recorded allegations of misconduct filed against an officer from 2000 to 2016. Allegations may originate from the public or from other officers in the department. For civilian allegations, the complainant must sign a sworn affidavit to certify that the allegation is true and correct. Only then is the complaint filed and investigated. We want to use complaints as a proxy measure of an officer's underlying "quality" since complaints are supposed to be reflective of officer misconduct.[12] Therefore,

---

Figure A1 for a map.

[10]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.

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

[12]Rozema and Schanzenbach 2018 find that bias-adjusted civilian allegations of misconduct can be used to predict officer misconduct that is likely to result in civil rights litigation and payouts. As a robustness check, we apply their empirical Bayes method to adjust officer complaints in our data, and find our results do not change.

we use all allegations about an officer, regardless of whether or not they were investigated.

Tactical Response Reports (TRR). Data on officer use of force come from 2004-2016 Tactical Response Reports. Officers are required to file Tactical Response Reports if they used any force while performing their duties. A TRR filing requirement can be triggered by three things: the subject's actions[13]; the officer's actions; or a subject who is injured or alleges injury resulting from the officer's use of force option. CPD publishes a Use of Force Model, which provides guidelines on the appropriate level of force to be used in response to a subject's actions and levels of resistance. The primary objective of using force is to ensure control of a subject with the reasonable force necessary based on the totality of the circumstances. Using the Use of Force Model as a guide, we classify officer force options into two broad categories of "weak use of force" and "strong use of force." Weak use of force includes force mitigation efforts, such as verbal direction and tactical positioning (which involve no physical touch), and control tactics, such as escort holds and wristlocks. Strong uses of force involve elevated levels of force that are generally intended to enact harm or injury on the subject.[14] The data only report use of force against adult persons. Appendix Table A2 outlines force options and our classification.

Sample restrictions. To construct a complete dataset on all officers in the Chicago Police Department, we require that officers receive a salary from DHR and appear in the unit assignment dataset. We focus on years 2009 to 2015 to maximize overlap across the different datasets. We further restrict our sample to officers at the rank of Police Officer who are always assigned to a geographic unit[15] and officer-supervisor relationships that lasted for

---

[13]One can categorize incidents involving subjects into three types: Cooperative Subject, Resister, and Assailant. A Cooperative Subject is one who is compliant without the use for physical force. There are two types of Resisters: a Passive Resister, who is a person who fails to comply (non-movement) with verbal or other direction; and an Active Resister, who is a person who attempts to create distance between himself or herself and the member's reach with the intent to avoid physical control and/or defeat the arrest. An Assailant is a subject who is using or imminently threatening the use of force against another person.

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

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

12 months. Our final analysis dataset has 6,518 Police Officers and 1,284 supervisors.

2.2   Crime Data

We use crime data from the Chicago Data Portal (https://data.cityofchicago.org), which contains reported incidents of crime that occurred in the City of Chicago since 2001. The dataset contains the primary type of crime, the date, location, and whether the crime led to an arrest. We construct monthly crime rates[16] by the geographic unit-level, separately for total crimes, property crimes, and violent crimes.[17] To capture a police unit's productivity, we also construct rates for crimes that led to an arrest.

2.3   Summary Statistics

This section provides descriptive statistics of Police Officers in our analysis sample. From Table 1, we see that most officers are male (73.7%) and white (46.4%), but blacks and Hispanics are also well-represented (23-27%). In fact, these three racial groups make up nearly 97% of our sample. The average CPD officer in our sample joined the force in 2000 at age 30. This indicates that at the start of our analysis dataset (2009), the average officer had been on the force for 9 years. There are no large differences between men and women in terms of age and start-year.

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

Table 2 presents racial and gender differences in various work measures. The first row is the probability of being nominated for an award in a particular month. For example, the average officer has a 2.3% chance of being nominated in a given month, which equates

---

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

[17]FBI does not record statistics for non-index crimes.

Table 1: Summary Statistics

	Police Officers			Supervisors		
	All	Male	Female	All	Male	Female
Race						
White	46.4%	48.6%	40.3%	69.7%	71.7%	60.7%
Black	26.8%	23.1%	37.3%	14.7%	12.4%	24.7%
Hispanic	23.2%	24.3%	20.2%	14.0%	14.0%	13.8%
Asian	3.1%	3.6%	1.8%	1.6%	1.8%	0.8%
Native American	0.4%	0.4%	0.4%	0.1%	0.1%	
Female	26.3%			19.2%		
Birthyear	1970.3	1970.7	1969.1	1965.3	1965.4	1965.1
Start Year	2000.0	2000.1	1999.6	1992.2	1992.2	1992.4
N	6,518	4,769	1,698	1,284	1,037	247

to about a 28% chance of being nominated in a given year. Whites and Hispanics have slightly higher than average likelihoods at 2.7% and 2.9%, respectively, while the likelihood for black officers is half the sample average (1.1%). The black-white difference is statistically significant at the 1% level. Similarly, the average female officer has a 1.2% chance—about the same odds as an average black officer—while the average male officer has a 2.7% chance of being nominated in a given month. The female-male difference is also statistically significant at the 1% level.

The second row in Table 2 lists the number of monthly complaints. The average officer receives about 0.04 complaints in a given month, equating to about 1 complaint every two years. This statistic is similar across race, but women receive slightly fewer complaints than men (0.03 vs 0.04). The female-male difference is statistically significant at the 1% level.

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

and complaints, arrest rates by female officers are similar to those by black officers (1.1 vs 1.2, respectively), while male officers are similar to white officers (2.1 vs 2.0, respectively). The female-male difference is statistically significant at the 1% level.

When comparing summary statistics for the different types of arrests, we see that the black-white difference in total arrests is driven by arrests for non-index crimes, which make up around 65% of all arrests. Here, the difference is about -0.70 arrests per month or -8.4 arrests per year and is statistically significant at the 1% level. Similarly, the female-male difference for non-index crimes is also large at -0.73 arrests per month.

It is important to note that arrests are not a comprehensive measure of policing quality and may be a biased measure. For example, a comparison of female and male officers' arrest records would suggest that female officers are less productive than male officers. However, a study by Miller and Segal 2018 finds that increasing the number of female police officers decreased the number of intimate partner homicides and increased the number of reports of domestic violence in the U.S. These outcome measures, which are important measures of social welfare, are not captured by arrests nor would they appear on an officer's record.

Another example is to consider drug and traffic arrests, which are presumably proactive in that they may indicate officers' initiative. We do not know whether an arrest stemmed from an incident that the officer initiated on his or her own authority, but we assume that drug and traffic arrests are more likely to have stemmed from officer-initiated traffic stops as compared to arrests for violent crimes. This classification of "proactive arrests", which allow for greater officer discretion, can also be seen as a delineation between appropriate and inappropriate uses of police authority.[18] In the last two rows of Table 2, we see that white officers are about 2.4-2.6 times more likely to arrest someone for drugs or traffic violations relative to black officers. Male officers are also about 1.8-2.8 times more likely than female

---

[18]We borrow this term from Worden et al. 2013, who analyzed the impact of a police agency's early intervention system, which aims at monitoring and managing police misconduct. The early intervention system was a training program for officers who exhibit patterns of problematic behavior, called the Officer-Civilian Interaction School, and researchers found that it lowered the number of proactive arrests with little impact on productivity.

officers to make these arrests. In contrast, the black-white difference for more serious crimes, like violent crimes, is economically small at -0.06 arrests a month and is weakly significant at 10%.

Further, given that a vast majority of arrestees is black[19] and the empirical evidence on racial bias in policing, we may be concerned that arrest numbers are an inaccurate measure of policing quality. That is, a negative black-white arrest gap may reflect that black officers are less productive than white officers (if arrests are a true measure of policing quality), or that white officers are of lower quality (if arrests are a biased measure of policing quality), or both. We sidestep this issue by assuming that supervisors when making nomination decisions care about the arrest and not whether or not it was motivated by bias.

Table 2: Racial and Gender Differences in Work Measures

Sample:	Everyone	White	Black	Hispanic	Male	Female
Nominated	0.0231	0.0272	0.0115***	0.0291	0.0271	0.0123***
Complaints	0.0425	0.0421	0.0422	0.0441	0.0467	0.0311***
Total Arrests	1.8246	2.0414	1.1939***	2.1641	2.0767	1.1308***
Violent	0.3672	0.3718	0.3117*	0.4201**	0.4010	0.2740***
Property	0.2713	0.2941	0.2015***	0.3034	0.2944	0.2078***
Non-Index	1.1861	1.3754	0.6807***	1.4406	1.3813	0.6489***
Drug	0.3125	0.3723	0.1411***	0.4050	0.3782	0.1315***
Traffic	0.1239	0.1471	0.0610***	0.1573	0.1411	0.0765***
Observations	250,872	111,876	70,572	59,148	184,020	66,852

Notes: This table lists monthly summary statistics for 6,518 police officers. Sample is at the officer-month level. Non-index arrests include arrests for non-property and non-violent crimes. Asteriks denote p-values from hypothesis tests of equal means between the minority group and the majority group. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3   Empirical Methodology

This section outlines the empirical methodology to examine whether majority police officers exhibit relative in-group bias in departmental award nominations. We focus on

---

[19]In our arrests data, about 70% of all arrestees are black.

awards that require a superior officer's nomination. Nominations may originate from any superior officer (i.e., sergeant), however, an analysis of the full sample of nominated officers is troublesome due to the decentralized nature of police work. In that case, we would only observe nominations from successful pairings between a Police Officer and a Sergeant, wherein the matching of officer race and sergeant race are endogenously determined.

To address this concern, we exploit an institutional feature of the Chicago Police Department. All sworn members of the Department are assigned a supervisor for annual performance evaluations.[20] Supervisors must be ranked higher than the supervisee, so they are also eligible to nominate their assigned officers for departmental awards.[21] We use the assignment to a new supervisor at the start of a calendar year to examine whether majority supervisors engage in in-group favoritism.[22] If officers are randomly assigned to supervisors, then we will be able to use the supervisor switch to approximate random assignment of a supervisor's race or sex to an officer. Thus, we can simply compare nomination rates when an officer was assigned to a black supervisor vs. a white supervisor, to test for the existence of relative in-group bias. This identification strategy is similar to Chetty, Friedman, and Rockoff 2014, which develops a quasi-experimental method by exploiting student exposure to teachers of varying value-added levels that is induced through teacher turnover. In this paper, we use officer exposure to supervisors of different races and sex that is induced through annual re-assignment of supervisors to officers.

Several facts suggest that officers do not game the supervisor assignment system. First, 88.6% of all officer-supervisor relationships are new relationships. Second, 78% of all supervisor relationships between 2009 and 2015 lasted one year. Although the vast majority of supervisor relationships last one year, because it is not a totality, we may be concerned that some officer-supervisor relationships may have been arranged outside of the random assign-

---

[20]Chicago Police Department, Career Development Directive, Employee Resource E05-01, retrieved from http://directives.chicagopolice.org/directives/data/a7a56e3d-12887ea9-ce512-887e-c3dce7cd73e28d57.html?hl=true

[21]In the data, all supervisors of Police Officers are at the rank of Sergeant.

[22]In the data, about 96 percent of officers are assigned to a supervisor in January of each calendar year.

ment system. Therefore, we restrict our analysis sample to all supervisor-officer relationships that last one year in order to minimize the number of endogenously formed supervisor relationships in our data. In the next section, we examine whether the data support the argument that officers are randomly assigned to supervisors.

3.1 Exogeneity of Officer Assignment by Productivity and Race/Sex

Throughout the paper, we want to interpret any change in nomination likelihood when minority officers are assigned to majority supervisors relative to when they are assigned to minority supervisors as a causal effect of supervisor race. The key assumption is that minority officers were not systematically assigned to majority officers in years when their productivity would have been particularly low for other reasons.[23] For example, if more productive white officers sort to white supervisors while more productive black officers sort to black supervisors, then we would see a negative black-white nomination gap among white supervisors and a positive black-white gap among black supervisors. This may appear to be in-group favoritism by both white and black supervisors but in reality it would be the result of sorting of high productivity police officers.

To test this, we examine whether any officer traits are statistically significantly correlated with supervisor race (sex) after conditioning on officer race (sex).[24] Because officers are assigned to supervisors at the Department level, we use the full officer sample rather than the analysis sample that is restricted to officers whose supervisor assignment lasted one year.

We estimate the following regression model separately by officer's group:

$$Maj_{it} = \beta_0 + W'_{it}\beta + X'_{it}\alpha + e_{it} \tag{1}$$

---

[23]Conversely, we would like to attribute any change in nomination likelihood when majority supervisors are assigned to minority officers vs. majority officers as a causal effect of officer race.

[24]In the data, officers are more likely to be assigned to a supervisor of the same race. The sorting concern, however, is not simply about white officers matching with white supervisors or black officers matching with black supervisors. It is about *more productive* officers matching with in-group supervisors. It is selection on officer's characteristics *and* race together that we care about.

Group is defined by race and sex, and we define $Ingroup$ to be the majority groups: white and male. Specifically, Maj_{it} is equal to 1 if officer i is assigned to a white or male supervisor in month t and 0 otherwise. W is a vector of baseline controls, such as birth year, race, tenure, unit fixed effects, year fixed effects, unit-year fixed effects, and an indicator for whether the officer and supervisor were in the same unit last month. X is a vector of officer characteristics, such as the number of officer arrests (total, non-index crimes, property crimes, and violent crimes), complaints, TRR filings, and uses of force (strong use of force and weak use of force). All of the variables in X are lagged by one month and include up to a cubic polynomial. Standard errors are clustered at the unit level.

To test our identification strategy, we are interested in whether officer characteristics, X, are predictive of an officer's assignment to an majority supervisor. These variables capture a sense of the officer's productivity and policing quality.

Table 3 presents estimates from the main control variables in X.[25] We are interested in whether more productive officers sort to white or male supervisors. Columns 1-3 present estimates for race, while columns 4-5 present estimates for gender. Because we are conducting multiple hypothesis tests to see which officer characteristics are statistically significant, we adjust the p-values using the Benjamini-Hochberg procedure. Asteriks denote coefficients whose adjusted p-value is less than 0.05.[26]

Most officers are assigned to white supervisors: about 78% for white officers, 61% for black officers, and about 71% for Hispanic officers. At the same time, there is little evidence that white, black, or Hispanic police officers sort to white supervisors based on their productivity measures. None of the characteristics in columns 1-3 are statistically significant. Column 4

---

[25]In addition, we estimate a version of equation (1) with and without the officer characteristics. The R-squared does not vary much between these two versions. It increases from 0.0923 for the baseline model to 0.1147 with the full set of controls when examining supervisor sorting by race, and from 0.0418 to 0.0585 when examining sorting by sex. That is, a flexible specification with officer productivity measures interacted with race/sex increases the predictive power by an additional 1-2 percentage-points. This suggests that officer characteristics do not explain much of the variation in supervisor race, which is reassuring.

[26]The issue with multiple testing is that there is typically a large probability that some of the true null hypotheses will be rejected. Therefore, some of the statistically significant estimates we see may not truly be statistically significant.

presents the evidence for male officers sorting to male supervisors, which is weak. Although arrests for non-index crimes are positively associated with being assigned to a male supervisor in the subsequent month (1.35 percentage-points), as discussed in Section 2.3 these are a less reliable measure of policing quality than arrests for violent or property crimes. We do not see evidence of positive sorting by female officers to male supervisors (column 5).

Table 3: Effect of Officer Characteristics on Supervisor Race and Sex

Dependent Variable:	Supervisor is White			Supervisor is Male	
Officer Sample:	White	Black	Hispanic	Male	Female
	(1)	(2)	(3)	(4)	(5)
Arrests					
Violent	0.00505	-0.0352	0.0200	-0.0151	0.0276
	(0.0164)	(0.0218)	(0.0157)	(0.0101)	(0.0158)
Property	0.0236	0.0280	8.20e-05	-0.000754	0.00262
	(0.0113)	(0.0285)	(0.0164)	(0.00979)	(0.0154)
Non-Index	-0.000457	-0.00268	-0.00226	0.0135***	0.00576
	(0.00570)	(0.0121)	(0.00765)	(0.00283)	(0.0108)
Complaints	-0.00767	-0.0785	-0.115	0.00958	-0.0584
	(0.0487)	(0.208)	(0.127)	(0.0350)	(0.110)
TRR Filings	-0.0442	-0.125	0.133	-0.149	-0.0171
	(0.113)	(0.192)	(0.192)	(0.0680)	(0.204)
Strong Force Ratio	0.0317	0.0558	0.108	0.0423	0.0349
	(0.0403)	(0.0778)	(0.0567)	(0.0306)	(0.103)
Weak Force Ratio	0.248	0.447	-0.0479	0.260	-0.239
	(0.213)	(0.290)	(0.193)	(0.119)	(0.409)
Same Unit	0.0131	-0.00760	0.00321	0.00814	-0.0473
	(0.0586)	(0.0640)	(0.0429)	(0.0244)	(0.0317)
Tenure	-0.00591	-0.0442***	-0.0161	0.0104	-0.00788
	(0.0122)	(0.0116)	(0.0211)	(0.00753)	(0.0170)
Observations	9,175	2,923	4,636	15,080	2,337
Baseline Mean	0.778	0.610	0.707	0.846	0.758
R-squared	0.121	0.178	0.164	0.060	0.138

Notes: Only estimates for main coefficients are shown. All variables except tenure are lagged by one month. Non-index crimes are composed of municipal code violations, traffic violations, warrants, drug-related crimes, and other. Standard errors are clustered at the unit level. Asterisks denote adjusted p-values using the Benjamini-Hochberg procedure for $\alpha = 0.05$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In summary, we do not find strong evidence that minority officers sort to majority super-

visors based on their productivity measures, conditional on their race and sex. This suggests that, in the absence of supervisor bias, we should not see any racial or gender differences in nominations. It also provides an indirect test of random assignment of officers to supervisors. If officers were randomly assigned to supervisors, conditional on race/sex, then we would not expect to see other traits correlated with supervisor assignment.

4   Results

In this section, we conduct two analyses to examine whether majority supervisors are biased against their minority officers. First, we test whether an officer's probability of nomination changes when the officer switches supervisors, and whether there are any differential changes if the officer is in the supervisor's in-group vs. out-group. Second, we examine whether an officer's arrest record, which we use as a proxy for productivity, affects the supervisor's likelihood of nomination, and whether there are any differential effects for minority officers.

4.1   Supervisor Switch

This analysis uses a difference-in-differences approach to test whether majority officers' nomination likelihood changes when they switch from a minority to a majority supervisor, and whether minority officers are differentially affected by a supervisor switch. Assuming equal productivity between minority and majority officers, there should be no difference in nominations in the absence of bias. Further, the DID estimates may be interpreted as causal if supervisors are randomly assigned to police officers. In other words, we use a supervisor switch to approximate random assignment of a supervisor's race or sex to an officer.

The regression sample is unique at the officer-switch level. A switch is defined as a change from one supervisor to a different supervisor. For example, say an officer has three different supervisors during the sample period. Then, this officer has two switches: Supervisor A to

Supervisor B, and Supervisor B to Supervisor C.[27] To examine whether majority supervisors are biased against minority officers, we focus on two types of switches: (1) from a non-white supervisor to a white supervisor, and (2) from a female supervisor to a male supervisor.

The regression model is:

$$Nom_{isj} = \beta_0 + \beta_1 \cdot PostSwitch_{isj} + (PostSwitch_{isj} \times N_i)'\delta + N_i'\gamma + X_{isj}'\alpha + e_{isj} \qquad (2)$$

where i denotes officer, s denotes switch, and j denotes supervisor. Nom_{isj} is equal to 1 if officer i was nominated for an award by supervisor j and 0 otherwise. $PostSwitch_{isj}$ is a binary variable equal to 1 if supervisor j is the new supervisor and equal to 0 if supervisor j is the old supervisor. N_i is a vector of binary indicator variables indicating the officer's group. When in-group is defined by race (e.g., when the sample is subsetted to non-white to white supervisor switches), N_i contains four binary indicator variables for whether the officer is Asian, black, Hispanic, or Native American. (White officers are the reference group.) When in-group is defined by sex (e.g., when the sample is subsetted to female to male supervisor switches), N_i is a female dummy variable.

X_{isj} is a vector of officer and unit characteristics during their assignment to supervisor j. Officer characteristics include officer tenure at the start of switch s, total complaints, and total arrests. Unit characteristics include the average crime rate, average violent crime rate, and average arrest rate. If the officer changed units during his or her assignment to supervisor j, then we use the modal unit. Standard errors clustered at the supervisor level.

The parameters of interest are β_1 and δ. β_1 tells us how nominations for majority (white or male) officers changed after they switched from a minority supervisor to a majority supervisor, while δ tells us whether nominations for minority officers are differentially affected, relative to majority officers, from the same supervisor switch. Table 4 column 1 presents estimates for a non-white to white supervisor switch, while column 2 present estimates for

---

[27]It is possible that an officer is assigned to Supervisor A in year 1, Supervisor B in year 2, and back to Supervisor A in year 3. These are defined as two unique switches: A-B and B-A.

a female to male supervisor switch.

White officers are 4.2 percentage-points more likely to be nominated after switching to a white supervisor, and this estimate is statistically significant at the 5% level. This is an effect size of 24% as the average white officer has a 17.5% chance of being nominated by his or her supervisor during the year. Black officers, however, do not benefit from the same switch. The black-white nomination gap widens by 3.5 percentage-points, which is nearly 20% of a white officer's baseline nomination probability. This estimate is significant at the 10% level. In contrast, the Hispanic-white nominations gap increases by 1.2 percentage-points (7%), though this estimate is not statistically significant.

Table 4: Effect of Supervisor Switch on Nomination Likelihood

| | Dependent Variable: Officer Nominated for Award | |
| Switch Type: | Non-White → White | Female → Male |
	(1)	(2)
Post-Switch	0.0420**	0.0502***
	(0.0183)	(0.0190)
Black Officer x Post-Switch	-0.0345*	
	(0.0208)	
Hispanic Officer x Post-Switch	0.0121	
	(0.0322)	
Female Officer x Post-Switch		-0.0235
		(0.0237)
Observations	4,706	3,350
Reference Group Mean	0.175	0.179
R-squared	0.517	0.537
Officer Characteristics	Yes	Yes
Unit Characteristics	Yes	Yes

Notes: Sample is at the officer-switch level. The reference group in column 1 is white officers and male officers in column 2. Officer characteristics include officer tenure at the start of switch, total complaints, and total arrests. Unit characteristics include the average crime rate, average violent crime rate, and average arrest rate. Standard errors are clustered at the supervisor level and are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We see similar patterns when we define in-group by gender (Table 4 column 2). The average male officer has a 17.9% chance of being nominated. Switching from a female

supervisor to a male supervisor is associated with a 5 percentage-point increase in nomination likelihood, or a 28% effect size. The female-male nomination gap widens by 2.4 percentage-points (13%), but it is not statistically significant.

Although the minority results are weakly significant at best, the large effect sizes are informative. They suggest that white supervisors do not treat all minority officers equally; there appears to be a bias against black officers in particular. There is weaker evidence that male supervisors are biased against female officers.

Last, we discuss the validity of our results. The identifying assumption for DID is that of parallel pre-trends. In our context, that means the minority nominations gap under minority supervisors should be zero, or that we have parallel leads.[28] Because our regression model includes officer fixed effects (equation 2), it cannot estimate the black-white gap or the female-male gap in the pre-switch period. However, we can manually estimate this through the Frisch-Waugh Theorem.[29] We estimate that the black-white difference is 0.002 and the female-male difference is essentially 0. Neither are statistically significant.

4.2 Nomination Likelihood by Arrest Record

To supplement our supervisor switch analysis, we exploit the panel nature of our data to examine the impact of officers' productivity on their likelihood of being nominated for an award. That is, conditional on the officer's arrest record, are there racial or gender differences in the probability of nomination? We estimate the following model, separately for white supervisors and male supervisors. The regression sample for this analysis is at the

---

[28]One cannot actually test for parallel pre-trends without data on the counterfactuals. However, one *can* empirically test for parallel leads.

[29]We do this separately to estimate the black-white nomination gap under non-white supervisors and the female-male nominations gap under female supervisors. First, we wipe the variation in nominations of all other covariates except officer race/sex (for race, we use a black dummy variable). Then we wipe the variation in the black/female dummy variable of the same covariates. Last, we regress the residuals from the first regression against the residuals from the second regression to obtain the black-white or female-male difference in nominations.

officer-month level.

$$Nom_{it} = \beta_0 + Arrests_{i,t-1}\beta + (Arrests_{i,t-1} \times N_i)'\delta + N_i'\gamma + X_{it}'\alpha + \tau_t + e_{it} \qquad (3)$$

where i denotes officer and t denotes month. Nom_{it} is equal to 1 if officer i was nominated for an award in month t and 0 if not. $Arrests_{i,t-1}$ is the number of arrests officer i made last month. Rather than including this as a continuous measure, we break up the number of arrests into six categories and include binary indicator variables for each: 0 arrests, 1, 2, 3, 4, and 5 or more arrests (the reference group is 0 arrests). We do this because the average number of monthly arrests is 2, and so we want to estimate impacts at lower numbers of arrests.

N_i is a vector of binary indicator variables indicating the officer's group. When in-group is defined by race (e.g., when the sample is subsetted to all officers assigned to white supervisors), N_i contains four binary indicator variables for whether the officer is Asian, black, Hispanic, or Native American. White officers are the reference group. When in-group is defined by sex (e.g., when the sample is subsetted to all officers assigned to male supervisors), N_i is a female dummy variable. $Arrests_{i,t-1} \times N_i$ are the interaction variables between the five arrest categories and the officer group categories.

X_{it} is a vector of officer and unit characteristics. These include officer fixed effects and time-varying characteristics, such as unit assignment, a cubic in tenure, and the number of complaints made against the officer. It also includes unit time-varying characteristics, such as crime rate, violent crime rate, and arrest rate. All time-varying variables are lagged by one month. We also include fixed effects for year and month, τ_t. Standard errors are clustered at the supervisor level.

The parameters of interest are β, which tell us how arrests affect the nomination likelihood for majority officers (white, male), and δ, which tells us the differential in likelihood between minority officers and their majority colleagues by number of arrests. We expect β to be

positive and increasing in the number of arrests. This is based on our belief that departmental awards are based on officer work productivity. If majority supervisors are not biased, then we would expect δ to be zero. A negative δ indicates that majority supervisors favor officers in their in-group (or, are biased against minority officers), whereas a positive δ indicates that majority supervisors favor officers not in their in-group.

Figure 1: Effect of Arrests on Nomination Likelihood

(a) White Supervisor   (b) Male Supervisor
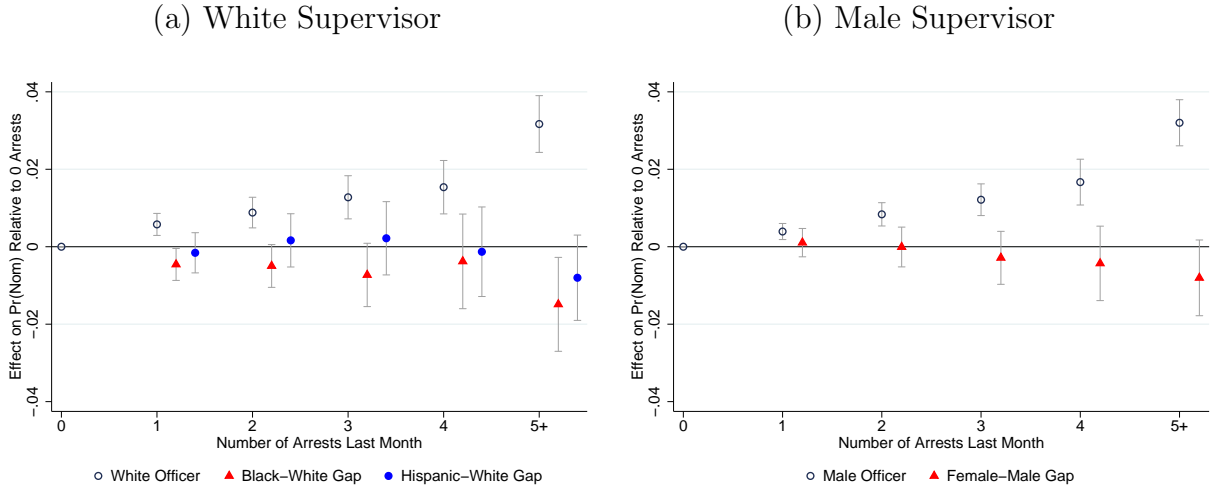
Notes: This graph depicts the impact of arrests on the probability of nomination, separately for white supervisors (Panel A) and male supervisors (Panel B). Sample is at the officer-month level. N is 176,552 for Panel A and 204,988 for Panel B. In Panel A, the reference group is white officers with no arrests. In Panel B, the reference group is male officers with no arrests. Estimates control for officer fixed effects, unit, tenure, complaints, unit-level crime rate, unit-level violent crime rate, and unit-level arrest rate, as well as year and month fixed effects. Standard errors are clustered at the supervisor level. Wings depict 95% confidence intervals.

Figure 1a presents estimates from equation (3). It illustrates how white supervisors' nomination behavior is affected by the officer's arrest record. There are increasing returns to having more arrests, with a marked increase for those with 5 or more arrests. The average white officer with zero arrests has a 1% chance of being nominated by his white supervisor for a departmental award the subsequent month. If the officer made one arrest, then the nomination likelihood increases by 0.58 percentage-points (57.5%). This estimate is statistically significant at the 5% level. If the officer made two arrests last month, which is the average number of monthly arrests for all officers, the nomination likelihood increases by 0.88

percentage-points (88%). This estimate is also statistically significant at the 5% level. For three arrests, the nomination likelihood increases by 1.3 percentage-points (127%), whereas the likelihood increases by 1.5 percentage-points (153%) for four arrests. Officers with 5 or more arrests, who are at the 84th percentile of monthly arrests, triple their probability of receiving a nomination (an increase of 3.2 percentage-points) relative to when they had zero arrests.

Black officers under white supervisors also see increasing returns to having more arrests, but the return is less. The black-white nomination gap widens by 0.46 percentage-points among officers with one arrest, relative to officers with zero arrests. The baseline black-white nomination gap for officers with no arrests is -0.9%, so this represents an effect size of 51%. Put another way, a black officer with the same arrest record, years of experience, and in the same unit as his white colleague, is 79.3% percent ($= -0.0046/0.0058$) less likely to be nominated for a departmental award by his white supervisor. This estimate is statistically significant at the 5% level.

The black-white nomination gap diminishes as the number of arrests increases, but does not go away. For example, a black officer with two or three arrests is 56-57% less likely to be nominated. This is statistically significant at the 10% level. For four arrests, it is -24.7% ($= -0.004/0.015$) though not statistically significant. For five or more arrests, it is -46.9% ($= -0.149/0.032$). This is statistically significant at the 5% level. It is important to interpret this disparity in the context of racial productivity differences. For example, black officers with 5 or more monthly arrests are at the 94th percentile of their distribution, while white officers are at the 81st percentile of their distribution. Yet, white supervisors are 47% less likely to nominate the black officer at the 94th percentile compared to the white officer at the 81st percentile.

In Figure 1a, we also examine whether white supervisors are less likely to nominate Hispanic officers, another racial minority in the Chicago Police Department. The magnitude of the Hispanic-white nomination gap, which ranges from 0.12 to 0.8 percentage-points, is

smaller than the black-white gap. Further, while there is a negative Hispanic-white gap for 1 arrest and 4+ arrests, the gap turns positive for 2-3 arrests. Last, none of the estimates are statistically significant. These results suggest that white supervisors do not exhibit similar behavior towards Hispanic officers and black officers, which is consistent with our findings from Table 4.

Figure 1b graph the same relationship for officers under male supervisors. As with Figure 1a, there are increasing returns to more arrests. The average male officer with zero arrests has a 1.1% probability of being nominated by his male supervisor. This is very close to the likelihood for the average white officer assigned to a white supervisor because almost half of all male officers are white. It is not surprisingly, therefore, that the pattern for majority officers is similar between Figure 1a and 1b. A male officer with one arrest is 0.39 percentage-points more likely (35.6%) to be nominated relative to when he had no arrests. This increases to 0.84 percentage-points (75.8%) with 2 arrests, 1.2 percentage-points (110%) with 3 arrests, 1.7 percentage-points (151%) with 4 arrests, and 3.2 percentage-points (290%) with 5 or more arrests. All of these estimates are statistically significant at the 5% level.

When we focus on female officers assigned to male supervisors, we find that women are 24-27% less likely than their male colleagues to be nominated, though the differences are not statistically significant.[30] Focusing on the right-tail of the officer distribution, however, sharpens the inequity. A female officer with 5 arrests is at the 96th percentile among women, while a male officer with 5 arrests is at the 81st percentile among men. However, the male supervisor is 25% less likely to nominate the woman than the man. This estimate is weakly significant with a p-value of 0.107.

---

[30]We check whether these results mask heterogeneity by race. We do not find strong evidence that either male supervisors or white male supervisors nominate white female officers at different rates from black female officers and Hispanic female officers.

5   Evidence on Mechanisms

In this section, we discuss several theories of discrimination that can explain our main findings. To streamline the analysis, we focus on white supervisors because we see evidence of in-group bias by white supervisors but not necessarily by male supervisors.

5.1   Animus

One explanation for why white supervisors are less likely to nominate black officers is that they may simply dislike black officers. We find support for this theory in Figure 1, where we find that white supervisors are less likely to nominate black officers *even conditional on the number of arrests*. A negative black-white nomination gap does not necessarily indicate animus; it may also be the result of statistical discrimination if supervisors believed that the average black officer is less productive than the average white officer. However, the black-white gap is still negative when we compare officers with the same number of arrests. Further, it persists when we focus on officers with 5 or more arrests–those who are clearly in the right-tail of the distribution. These results suggest that a mechanism other than statistical discrimination is at play.

At the same time, animus is difficult to determine without a scientific measure like the Implicit Association Test. We attempt to detect animus using two supervisor-specific measures: the black-white nomination gap and the black-white TRR (use of force) gap. These measures on their own may not be enough to indicate animus. For example, a negative black-white nomination gap may be due to animus or statistical discrimination. A positive black-white gap in TRR filings indicates that an officer is more likely to use force against black subjects relative to white subjects. This may reflect animus, but it may also reflect unobservable racial differences in subject behavior. If, for example, black subjects tend to be more aggressive than white subjects on average, then this may explain why officers use more force with black subjects than white subjects.

In contrast, considering both of these measures, which are constructed from interactions with two different populations, may help to identify animus.[31] If there is racial animus, then we may expect there to be a negative relationship between a supervisor's nomination record towards black officers and his use of force on black subjects. That is, a supervisor with animus would be less likely to nominate a black officer relative to a white officer *and* be more likely to use force against a black subject relative to white subject.

Using TRR data from 2004 to 2015, we calculate the black-white difference in TRR filings for each officer conditional on the officer having filed at least one TRR during this time period.[32] If an officer filed no TRRs between 2004 and 2015, then the officer is not in our sample.[33] The black-white nomination gap is calculated as the difference between the number of times the supervisor nominated a black officer and the number of times supervisor nominated a white officer, divided by the total number of assigned officers. Only supervisors who have both black and white assigned officers are included in this exercise.

Figure 2 plots these two measures for the 569 supervisors in our analysis sample for whom we have TRR information. For white supervisors, the plot shows a downward sloping line of best fit, indicating that white officers who are less likely to nominate black officers are also more likely to use force against black subjects. The coefficient on the best fit line is -1.17, meaning that a supervisor who nominates all his white officers and none of his black officers also has 1.17 more TRR filings against black subjects, relative to a supervisor who nominates black and white officers at equal rates. This represents a 31% increase in the use of force against black subjects as the average white supervisor in this analysis sample[34] has

---

[31]We define animus to be towards aggregate racial groups. For example, we assume that animus towards blacks means animus towards all blacks, not animus towards black subjects but not towards black colleagues.

[32]The black-white TRR gap is defined as the number of TRR filings where the subject was black less the number of TRR filings where the subject was white. Missings are treated as zeroes. So, for example, if Officer X had filed 4 TRRs where the subject was black but no TRRs where the subject was white, then Officer X's black-white TRR gap is equal to 4.

[33]This is why we were able to match 569 out of the 1,284 supervisors in our analysis sample. Including these officers does not affect our results; we gain an additional 105 white supervisors and the correlation coefficient changes from -1.17 to -1.30. Statistical significance remains at 1%.

[34]Recall that the sample for this exercise is restricted to supervisors who are assigned to both black and white officers and have filed at least one TRR between 2004 and 2015.

Figure 2: Nominations vs TRR Filings by Supervisor Race
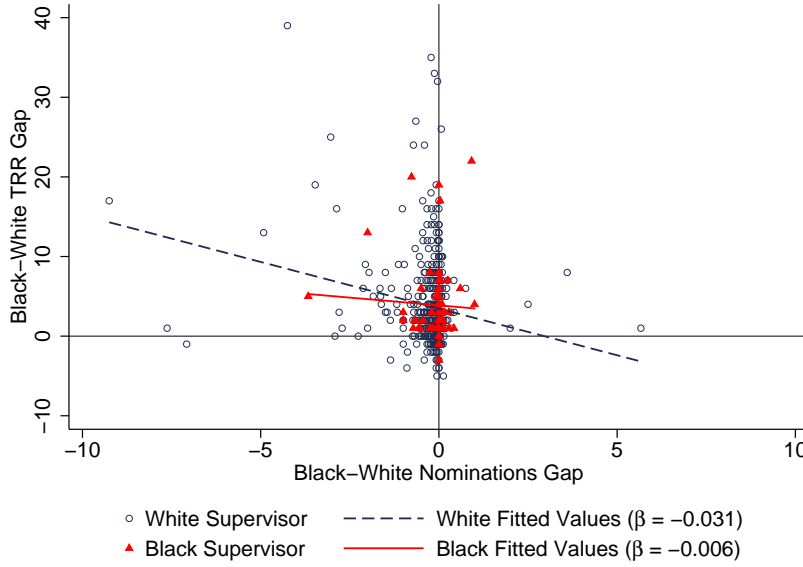
Notes: This graph plots the black-white nomination gap on the x-axis against the black-white TRR (use of force) gap on the y-axis for each supervisor. N = 569. The black-white nomination gap is calculated as the difference between the number of times the supervisor nominated a black officer and the number of times supervisor nominated a white officer, divided by the total number of assigned officers. The black-white TRR gap is defined as the number of TRR filings where the subject was black less the number of TRR filings where the subject was white.

a black-white TRR gap of 3.75. This estimate is statistically significant at the 1% level. For black supervisors, the slope coefficient is -0.39 and not statistically significant, suggesting that use of force is independent of their nomination decisions. These findings are consistent with a theory of animus.

5.2 Less Interaction

White supervisors may be less likely to nominate black officers because they are less likely to interact with them and, therefore, are less likely to be informed of their accomplishments (Glover, Pallais, and Pariente 2017). To test this theory, we exploit an institutional feature that randomizes the quarter in which officers are evaluated by their supervisor.

All supervisors are required to conduct annual evaluations of their assigned officers, and

28

this evaluation must take place during the quarter prior to the quarter in which the officer joined the Department. For example, if an officer started his career in July (Q3), then his annual evaluation must take place in the second quarter of every calendar year. Because start dates are randomly determined by a lottery number, this means that the evaluation quarter is essentially randomly assigned across officers.[35]

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

$$Nom_{it} = \beta_0 + \sum_{\substack{q=-3, \\ \neq -1}}^{3} \mathbb{1}\{RQ = q\}\delta^q + \left(black_i \times \sum_{\substack{q=-3, \\ \neq -1}}^{3} \mathbb{1}\{RQ = q\} \right) \beta_1^q + X_{it}'\alpha + \tau_t + e_{it} \quad (4)$$

where i denotes officer and t denotes month. Nom_{it} is equal to 1 if officer i was nominated for an award in month t and 0 if not. $black_i$ is a dummy variable equal to 1 if officer i is black and 0 if the officer is white. X contains controls for officer characteristics (officer fixed effects, tenure, unit, complaints, arrests) and unit characteristics (crime rate, violent crime rate, arrest rate). τ_t includes month and year fixed effects. Standard errors are clustered at the supervisor level.

The second term is a set of binary indicator variables for each quarter relative to the evaluation quarter ($RQ = 0$). The reference quarter is the quarter before the evaluation quarter, $RQ = -1$. The coefficients δ^q tell us how nomination likelihoods for white officers change across relative quarters. If an interaction effect exists, we expect it to be enhanced

---

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

in the evaluation quarter (δ^0), when supervisors are required to evaluate their officers.

The third term is an interaction between the black indicator variable and the relative quarter indicator variables. The coefficients β_1^q depict how the black-white nomination gap evolves relative to $RQ = -1$. If white supervisors are equally likely to nominate their black and white officers, then we expect β_1^q to be zero.

Figure 3: Probability of being Nominated by Quarter
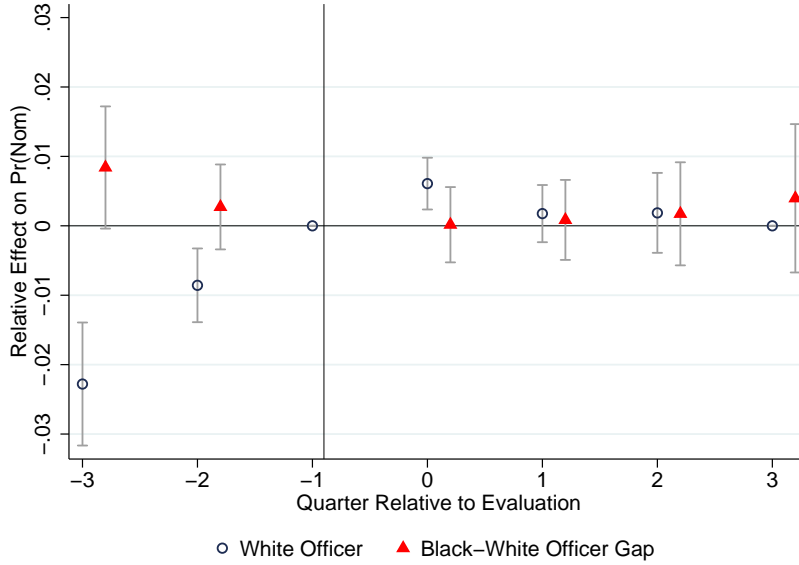
Notes: This graph depicts the impact of arrests on the probability of nomination by white supervisors. The reference group is white officers in $RQ = -1$. Estimates control for officer fixed effects, unit, tenure, arrests, complaints, unit-level crime rate, unit-level violent crime rate, and unit-level arrest rate. Standard errors are clustered at the supervisor level. Wings depict 95% confidence intervals.

Figure 3 plots our estimates for δ^q and β_1^q. In the quarter before their evaluation, white officers have a 3.5% chance of being nominated by their white supervisor. This increases by 0.6 percentage-points or 17.6% in the evaluation quarter. This increase occurs for both white and black officers assigned to white supervisors; the black-white gap in the evaluation quarter is 0.002 percentage-points larger relative to the one in the previous quarter and not statistically significant. In subsequent quarters, the nomination likelihood essentially returns to its initial value. Further, the black-white difference remains negligent.

These results suggest an interaction channel exists, and that it disappears in subsequent quarters. This pattern is consistent with what we know about the nature of police work; assigned supervisors are not directly managing police officers on the job and, therefore, may not always be aware of their work accomplishments.

We would like to discuss the large and negative estimate in $RQ = -3$ for white officers. Recall that new supervisors are assigned in January of every year. Therefore, this estimate is coming from officers whose evaluation quarter is Q4 or the last three months of the year (see Table A3). This means that, in addition to any interaction effect from the required evaluation, we are capturing the fact that the officer's annual evaluation occurs at the end of the calendar year. Relative to other officers whose evaluations occur earlier in the calendar, these "Q4 officers" are less likely to interact with their supervisors early in the year, and therefore their relative increase is larger. Put another way, there is only one quarter for which $RQ = -3$: Q1. As there is no other variation, we are unable to separate the impact of calendar quarter from the interaction effect. However, our results are robust to dropping these officers.

Second, why is the black-white gap less negative in $RQ = -3$? The gap, at 0.008 percentage-points, is approximately half of the white officer's estimate of -2.3 percentage-points. This is because the relative increase throughout the year is not as large for black officers as it is for white officers. Figure A2 plots the mean nomination probability by race and evaluation quarter. The probability for black officers increases by 1 percentage-point between $RQ = -3$ and $RQ = 0$; the percentage-point increase for white officers is double that.

To summarize, we find evidence in support of an interaction channel. Nominations increase the most in the evaluation quarter, suggesting that white supervisors do not interact regularly with their assigned officers until they are required to do so. This behavior, however, is towards both white and black officers, limiting its ability to explain the black-white nomination gap.

5.3 Collective Threat

The preceding analysis explored theories that can directly explain why black officers are nominated less frequently than white officers. Here, we explore a channel through which negative stereotypes may indirectly affect the nomination gap.

Studies have shown that negative stereotypes, which can adversely affect one's own performance, can result in minority achievement gaps (Coffman 2014; Glover, Pallais, and Pariente 2017). For example, Steele and Aronson 1995 find that black students perform worse on the verbal GRE exam when they are told that the test is supposed to diagnose intellectual ability compared to when they are told that the test has nothing to do with ability. These privately held beliefs can have negative externalities. For example, Eble and Hu 2019 find that parental beliefs about gender differences in math ability is transmitted to their children. Carlana 2019 finds that teacher stereotypes about gender-science ability lead girls to under-perform in math and to self-select into less demanding high schools.

An important implication of negative stereotypes in the workplace is the manifestation of collective threat, or the concern by a minority group member of being devalued by an in-group member's behavior (Cohen and Garcia 2005). In this situation, minority supervisors may perseverate labor market disparities rather than reduce them. For example, Srivastava and Sherman 2015 find that female managers do not reduce, but rather widen, the gender wage gap among their lower-performing subordinates.[36] In the criminal justice system, Depew, Eren, and Mocan 2017 find that black and white juveniles randomly assigned to same-race judges are more likely to get incarcerated and receive longer sentences.

We examine whether or not black supervisors experience collective threat. Our approach is motivated by the idea that collective threat may generate differences in quality between black and white officers who are nominated by black supervisors. We compare the arrest

---

[36]The authors examine longitudinal personnel records at a firm where managers have discretion over employee salaries and find that when low-performing women switch from a high-performing male supervisor to a high-performing female supervisor, they have lower salaries than low-performing men who made the same switch.

record of all officers nominated for a departmental award. Specifically, we look at data from month $t - 1$ for all officers who were nominated in month t. If black supervisors experience collective threat, then we would expect to see black officers held to a higher standard than white officers when nominated by black supervisors, but not necessarily when nominated by white supervisors.

The regression model is as follows:

$$Arrests_{it} = \beta_0 + N_i'\gamma + \beta_2 WhiteSup_{it} + (WhiteSup_{it} \times N_i)'\delta^r + X_{it}'\alpha + \tau_t + e_{it} \quad (5)$$

where i denotes officer and t denotes month. $Arrests_{it}$ is officer i's arrests in month t in standard deviations. We calculate a race-specific standard deviation for each officer.

N_i is a vector of binary indicator variables indicating the officer's race. It contains four binary indicator variables for whether the officer is Asian, Hispanic, Native American, or white. Black officers are the reference group. We are interested in the coefficient on the white officer dummy variable, γ^W; this will tell us the white-black difference in total arrests conditional on being nominated by a black supervisor. A negative estimate indicates that the average black officer nominated by a black supervisor is more productive—using arrests as a measure of productivity—than the average white officer nominated by a black supervisor.

$WhiteSup_{it}$ is an indicator variable equal to 1 if officer i was assigned to a white supervisor in month t and equal to 0 if assigned to a black supervisor. We are also interested in the coefficient on $WhiteSup_{it}$, β_2. This will tell us whether white and black supervisors have different thresholds for nominating black officers. A negative estimate indicates that the average black officer nominated by a black supervisor is more productive (again using arrests as a proxy for productivity) than the average black officer nominated by a white supervisor.

$WhiteSup_{it} \times N_i$ are the interaction variables between supervisor's race and the officer's race. X contains controls for officer tenure and unit characteristics (unit assignment, unit-level crime rate, unit-level violent crime rate, unit-level arrest rate). τ_t includes month and

year fixed effects. Standard errors are clustered at the supervisor level.

Table 5 reports estimates for γ^W and β_2. Columns 1 and 2 report differences in violent crime arrests and property crime arrests—more serious offenses that the Federal Bureau of Investigation classifies as "index crimes".[37] Column 3 reports differences in arrests for non-index crimes, which are less serious offenses such as gambling, vandalism, and disorderly conduct. Although most of the estimates are not statistically significant, the signs and magnitudes are consistent with the story that black supervisors experience collective threat.

The average black officer nominated by a black supervisor is 0.37 standard deviations above the black mean, or at the 64th percentile, in terms of violent crime arrests. In contrast, the average white officer is 0.19 standard deviations above the white mean, or at the 57th percentile. For the property crime arrests, the disparity is larger. The average black supervisor nominated a black officer at the 67th percentile among all black officers (0.45 standard deviations), but a white officer at the 56th percentile among all white officers (0.16 standard deviations). This difference is statistically significant at the 5% level. For non-index crime arrests, the white-black officer difference is negative at -0.33, again indicating that black officers are at a higher percentile relative to white officers.

The second row details how thresholds for black officers differ by supervisor race. These estimates are again all negative, suggesting that black supervisors have a higher standard than white supervisors. However, none of the estimates are statistically significant. In terms of violent crime arrests, the average black officer nominated by a white supervisor is 0.11 standard deviations below the average black officer nominated by a black supervisor. We see similar differences for property crime arrests (-0.11 standard deviations) and non-index crime arrests (-0.16 standard deviations). These results are also consistent with the theory of collective threat, in that black supervisors hold their in-group members to a higher standard relative to white supervisors.

---

[37]This definition is taken from the Federal Bureau of Investigation's National Incident-Based Reporting System.

Table 5: Racial Differences in Arrests Among Nominated Officers

| | Dependent Variable: Number of Arrests (std dev) | | |
| | Violent | Property | Non-Index |
	(1)	(2)	(3)
White Officer	-0.180	-0.294**	-0.328
	(0.154)	(0.121)	(0.410)
White Supervisor	-0.112	-0.106	-0.159
	(0.139)	(0.159)	(0.422)
Observations	13,358	13,358	13,358
Black Officer Mean	0.371	0.449	1.386
R-squared	0.039	0.031	0.136
Month and Year FE	Yes	Yes	Yes
Officer Tenure	Yes	Yes	Yes
Unit Characteristics	Yes	Yes	Yes

Notes: Sample is restricted to all officers in the month before the month that they were nominated. The reference group is black officers assiged to black supervisors. Unit characteristics include crime rate, violent crime rate, and arrest rate. Standard errors are clustered at the supervisor level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

5.4 Interpretation

In this section, we examine the extent to which different factors can explain the minority nomination gap. How much of the gap can be explained by officer productivity vs. supervisor interaction vs. animus? The sample for this exercise is at the officer-month level, restricted to all officers who are assigned to a majority supervisor. When focusing on race, the majority supervisor is white. When focusing on gender, the majority supervisor is male. The regression model is a variant of equation (4), but without the interaction between relative quarter and officer race because we did not find significant differential effects by officer

race.[38]. It is as follows:

$$Nom_{it} = \beta_0 + N_i\gamma + \sum_{\substack{q=-3, \\ \neq 0}}^{3} \mathbb{1}\{RQ = q\}\delta^q + X_{it}'\alpha + \tau_t + e_{it} \qquad (6)$$

where i denotes officer and t denotes month. Nom_{it} is equal to 1 if officer i was nominated for an award in month t and 0 if not. X_{it} is a vector of officer and unit characteristics. Officer characteristics include a cubic in tenure, the number of complaints made against the officer, a cubic in total arrests, violent crime arrests, property crime arrests, and non-index crime arrests. Unit characteristics include unit assignment, unit-level crime rate, unit-level violent crime rate, and unit-level arrest rate. τ_t includes fixed effects for year and month. Standard errors are clustered at the supervisor level.

RQ is a set of binary indicator variables for each quarter relative to the evaluation quarter. We use these to proxy for the information channel, and set the reference quarter as the evaluation quarter ($RQ = 0$) because that is the quarter in which supervisors are required to interact with their assigned officers.

N_i is a vector of binary indicator variables indicating the officer's group. When in-group is defined by race (e.g., when the sample is subsetted to all officers assigned to white supervisors), N_i contains four binary indicator variables for whether the officer is Asian, black, Hispanic, or Native American. White officers are the reference group. When in-group is defined by sex (e.g., when the sample is subsetted to all officers assigned to male supervisors), N_i is a female dummy variable. The parameters of interest are the coefficient on the black indicator variable in N_i, γ^B, and the coefficient on the female indicator variable in N_i, γ^F. These will tell us, respectively, how the black-white nomination gap and the female-male nominations gap evolve as we include additional controls.

Table 6 Panel A presents estimates for the black-white nomination gap. We begin with the unconditional gap in column 1, which states that white supervisors are 1.8 percentage-

---

[38] As a robustness check, we estimate a model where we include interactions and find that the estimates do not change much. The black-white gap is remains around -0.9 or -0.10 percentage-points.

points (57%) less likely to nominate black officers relative to white officers. In column 2, this estimate reduces by about half, from 1.8 to nearly 0.9 percentage-points, when we control for time and officer characteristics. In column 3, we include unit characteristics, such as crime rate, and find that the estimate increases slightly to 0.97 percentage-points. This is because black officers are more likely to be assigned to units with higher crime rates, and nomination rates are positively correlated with higher crime rates. In column 4, we include indicator variables for the quarter relative to the evaluation quarter, in order to control for interactions between supervisors and officers. This does not change the estimate much; the black-white gap remains at around -0.97 percentage-points. This is consistent with our results in Section 5.2, which did not find differential effects by officer race and evaluation quarter. All of the estimates are statistically significant at the 1% level.

Panel B repeats this same exercise for the female-male nominations gap. The unconditional gap is 1.7 percentage-points, or 52%. That is, male supervisors are 52% less likely to nominate a female officer than a male officer. If we control for officer productivity, this gap reduces by about 45%. So, even after accounting for the number of arrests and complaints, male supervisors are 25% less likely to nominate a female officer. These estimates are statistically significant at the 1% level.

In column 3, we control for neighborhood characteristics in case female officers may be differentially assigned to less dangerous neighborhoods, thus providing an explanation for the female-male nominations gap. However, this has little effect on the gap, which decreases to 0.75 percentage-points from 0.8 percentage-points. In column 4, we control for supervisor interaction, and the estimate, again, changes little (0.73 percentage-points).

To summarize, our findings suggest that officer productivity, neighborhood characteristics, and supervisor interaction can explain about half of the minority nominations gap. For the black-white gap, they can explain about 45%, while they can explain 56% of the female-male gap. However, in both cases roughly half of the original gap remains. This may be attributed to animus, which we found support for in Section 5.1.

Table 6: Minority Nominations Gap

| | Dependent Variable: Officer Nominated for Award | | | |
	(1)	(2)	(3)	(4)
Panel A: Black-White Gap				
Black	-0.0177***	-0.00898***	-0.00969***	-0.00973***
	(0.00187)	(0.00150)	(0.00212)	(0.00212)
Observations	176,552	176,552	176,552	176,552
White Officer Mean	0.031	0.031	0.031	0.031
R-squared	0.002	0.035	0.037	0.039
Panel B: Female-Male Gap				
Female	-0.0166***	-0.00801***	-0.00749***	-0.00733***
	(0.00141)	(0.00107)	(0.00109)	(0.00108)
Observations	204,988	204,988	204,988	204,988
Male Officer Mean	0.032	0.032	0.032	0.032
R-squared	0.002	0.037	0.038	0.039
Month and Year FE		Yes	Yes	Yes
Officer Characteristics		Yes	Yes	Yes
Unit Characteristics			Yes	Yes
Evaluation Quarter				Yes

Note: Panel A sample is restricted to all officers assigned to white supervisors, while Panel B sample is restricted to all officers assigned to male supervisors. Officer characteristics include tenure, complaints, and arrests. Unit characteristics include unit fixed effects, crime rate, violent crime rate, and arrest rate. Evaluation quarter controls are included as a triple interaction between officer group indicator variables (race or gender), relative quarter indicator variables, and month indicator variables. Standard errors are clustered at the supervisor level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6 Conclusion

Racial bias has been extensively documented in a variety of settings, including hiring decisions (Bertrand and Mullainathan 2004), sports umpires (Parsons et al. 2011), jury and judicial decisions (Park 2017; Flanagan 2018), and bail decisions (Arnold, Dobbie, and Yang 2018). The increasing availability of police administrative data has allowed researchers to carefully examine and detect bias in police interactions as well. A potential solution that has been put forth is to increase racial and gender diversity among officers. Sklansky 2005

identifies three mechanisms through which diversity may affect policing: minority officers may have a distinctive set of abilities and skills ("competency effects"); demographic diversity of a police department may affects its relations with the community ("community effects"); and workforce diversity may affect the internal dynamics of the department ("organizational effects").

Past studies have touched upon the first two channels. For example, Miller and Segal 2018 finds that the integration of women in the US police force led to increased reports of violent crime and domestic violence against women, but had no impact on reporting for male victims. Relatedly, Anwar, Bayer, and Hjalmarsson 2012 finds that the inclusion of women on jury panels led to an increase in convictions for sex offenses. McCrary 2007 finds that court-ordered racial hiring quotas for police departments during the 1970s were targeted at big cities with growing black populations.

This paper sheds light on the third mechanism. To what extent would a diversity initiative be effective at fighting negative stereotypes? This question is of particular relevance to law enforcement, where minorities are less represented at higher ranks. For example, white males comprised 40% of all entry-level police officers in the Chicago Police Department in 2015, and 56% of those at the rank of Sergeant or higher.[39]

Exploiting quasi-random variation in supervisor assignment, we find that white supervisors are less likely to nominate black officers than white or Hispanic officers. We find weaker evidence that male supervisors are less likely to nominate female officers than male officers. These results control for officer characteristics, including arrest record and misconduct. Further, we find a significant, negative black-white nomination gap even among officers who are the right-tail of the distribution. We explore several theories of discrimination that can explain our main findings, and the evidence is most consistent with a theory of animus against black officers. These findings suggest that simply increasing the diversity of incoming recruits may be ineffective.

---

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

References

Antonovics, Kate, and Brian 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, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The impact of jury race in criminal trials." *The Quarterly Journal of Economics* 127 (2): 1017–1055.

Anwar, Shamena, and Hanming Fang. 2006. "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence." *American Economic Review* 96 (1): 127–151.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial bias in bail decisions." *The Quarterly Journal of Economics* 133 (4): 1885–1932.

Beaman, Lori, Niall Keleher, and Jeremy Magruder. 2018. "Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi." *Journal of Labor Economics* 36 (1): 121–157.

Bertrand, Marianne, and Sendhil 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 Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope. 2019. "Inaccurate Statistical Discrimination." Technical Report, National Bureau of Economic Research.

Carlana, Michela. 2019. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *The Quarterly Journal of Economics.*

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American Economic Review* 104 (9): 2593–2632.

Coffman, Katherine Baldiga. 2014. "Evidence on self-stereotyping and the contribution of ideas." *The Quarterly Journal of Economics* 129 (4): 1625–1660.

Cohen, Geoffrey L, and Julio Garcia. 2005. ""I Am Us": Negative Stereotypes as Collective Threats." *Journal of Personality and Social Psychology* 89 (4): 566.

Coviello, Decio, and Nicola Persico. 2015. "An Economic Analysis of Black-White Disparities in the New York Police Departments Stop-and-Frisk Program." *The Journal of Legal Studies* 44 (2): 315–360.

Depew, Briggs, Ozkan Eren, and Naci Mocan. 2017. "Judges, juveniles, and in-group bias." *The Journal of Law and Economics* 60 (2): 209–239.

Donohue III, John J, and Steven D Levitt. 2001. "The impact of race on policing and arrests." *The Journal of Law and Economics* 44 (2): 367–394.

Eble, Alex, and Feng Hu. 2019. "How important are beliefs about gender differences in math ability? Transmission across generations and impacts on child outcomes." Technical Report.

Egan, Mark L, Gregor Matvos, and Amit Seru. 2018. "When Harry fired Sally: The Double Standard in Punishing Misconduct." Technical Report, National Bureau of Economic Research.

Flanagan, Francis X. 2018. "Race, gender, and juries: Evidence from north carolina." *The Journal of Law and Economics* 61 (2): 189–214.

Fryer, Roland G. 2018. "An Empirical Analysis of Racial Differences in Police Use of Force." *Journal of Political Economy*, vol. Forthcoming.

Giuliano, Laura, David I Levine, and Jonathan Leonard. 2009. "Manager race and the race of new hires." *Journal of Labor Economics* 27 (4): 589–631.

Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores." *The Quarterly Journal of Economics* 132 (3): 1219–1260.

Goncalves, Felipe, and Steven Mello. 2018. "A Few Bad Apples?: Racial Bias in Policing." Technical Report.

Hengel, Erin. 2019. "Publishing while Female. Are women held to higher standards? Evidence from peer review."

Hoekstra, Mark, and CarlyWill Sloan. 2019. "The Effect of Police Officer Race on Use of Force." Technical Report.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial bias in motor vehicle searches: Theory and evidence." *Journal of Political Economy* 109 (1): 203–229.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2019. "The Bias is Built In: How Administrative Records Mask Racially Biased Policing." Technical Report.

Kofoed, Michael, and Elizabeth mcGovney. 2019. "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice." *Journal of Human Resources*, vol. Forthcoming.

Langan, Andrew. 2018. "Female Managers and Gender Disparities: The Case of Academic Department Chairs." Technical Report.

McCrary, Justin. 2007. "The Effect of Court-ordered Hiring Quotas on the Composition and Quality of Police." *American Economic Review* 97 (1): 318–353.

Miller, Amalia, and Carmit Segal. 2018. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *Review of Economic Studies*, vol. Accepted.

Nix, Justin, Bradley A Campbell, Edward H Byers, and Geoffrey P Alpert. 2017. "A Bird's Eye View of Civilians Killed by Police in 2015." *Criminology and Public Policy* 16 (1): 309–340.

Park, Kyung H. 2017. "The impact of judicial elections in the sentencing of black crime." *Journal of Human Resources* 52 (4): 998–1031.

Parsons, Christopher A, Johan Sulaeman, Michael C Yates, and Daniel S Hamermesh. 2011. "Strike three: Discrimination, incentives, and evaluation." *American Economic Review* 101 (4): 1410–35.

Persico, Nicola. 2002. "Racial profiling, fairness, and effectiveness of policing." *American Economic Review* 92 (5): 1472–1497.

Rozema, Kyle, and Max Schanzenbach. 2018. "Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct." *American Economic Journal: Economic Policy.*

Sarsons, Heather. 2018. "Interpreting signals in the labor market: evidence from medical referrals." Technical Report.

———. 2019. "Gender differences in recognition for group work." *Journal of Political Economy*, vol. Accepted.

Sklansky, David Alan. 2005. "Not Your Father's Police Department: Making sense of the new demographics of law enforcement." *J. Crim. L. & Criminology* 96 (3): 1209–1243.

———. 2011. "The persistent pull of police professionalism." *New Perspectives in Policing, March.*

Srivastava, Sameer B, and Eliot L Sherman. 2015. "Agents of change or cogs in the machine? Reexamining the influence of female managers on the gender wage gap." *American Journal of Sociology* 120 (6): 1778–1808.

Steele, Claude M, and Joshua Aronson. 1995. "Stereotype threat and the intellectual test performance of African Americans." *Journal of personality and social psychology* 69 (5): 797.

United States Department of Justice. March 4, 2015. "Investigation of the Ferguson Police Department." Technical Report, Civil Rights Office.

Weisburst, Emily. 2018. "Whose Help Is on the Way? The Importance of Individual Police Officers in Law Enforcement." Technical Report.

West, Jeremy. 2018. "Racial Bias in Police Investigations." Technical Report.

Worden, Robert E, Moonsun Kim, Christopher J Harris, Mary Anne Pratte, Shelagh E Dorn, and Shelley S Hyland. 2013. "Intervention with problem officers: An outcome evaluation of an EIS intervention." *Criminal justice and behavior* 40 (4): 409–437.

Table A1: Department Awards

1	*Superintendent's Award of Valor* 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.
2	*Superintendent's Award of Merit* 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.
3	*Police Blue Star Award* is granted to any sworn member who has been seriously, critically, or fatally injured while in the performance of police duty.
4	*Police Blue Shield Award* 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.
5	*Superintendent's Award of Tactical Excellence* 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.
6	*Arnold Mireles Special Partnership Award* for making a significant impact upon the quality of life within their community by identifying and resolving problems.
7	*Special Commendation* for making a significant impact on public safety or crime prevention.
8	*Lifesaving Award* for a successful effort in saving a human life that involved exceptional courage or performance.
9	*Police Officer of the Month* for performance of duty during a specific month was characterized by such exceptional professional skill that it merits recognition by the entire Department.
10	*Chicago Police Leadership Award* for exemplary service, dedication, and leadership.
11	*Department Commendation* 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.
12	*Problem Solving Award* for an exemplary effort to identify, analyze, and successfully respond to causes, conditions, and problems that may lead to crime and neighborhood disorder.
13	*Joint Operations Award* 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.
14	*Unit Meritorious Performance Award* for exhibiting exceptional professional skill and conduct during a coordinated action.
15	*Traffic Stop of the Month Award* for excellence in conducting professional traffic stops that result in quality arrests.
16	*Top Gun Arrest Award* for exceptional commitment to the recovery of illegal firearms.
17	*Special Service Award* for contributing to any event that has a significant impact upon the historical direction and operations of the Department.
18	*Honorable Mention Certificate* for demonstrating outstanding performance above and beyond that required by the member's assignment.

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

Table A2: Use of Force Options and Member Response

Member's Response	Our Classification

Force Mitigation Efforts
  Member Presence
  Zone of Safety
  Verbal Direction/Control Techniques
  Movement to Avoid Attack
  Specialized Units
  Tactical Positioning
  Additional Unit Members
  None
  Other
	Mitigation

Control Tactics
  Escort Holds
  Wristlock
  Armbar
  Control Instrument
  Pressure Sensitive Areas
  Emergency Handcuffing
  Other
	Control tactics

Response without Weapons
  Open Hand Strike
  Take down
  Elbow strike
  Close hand strike/Punch
  Knee strike
  Kicks
  Other
	No Weapon

Response with Weapons
  OC/Chemical Weapon
  OC/Chemical Weapon w/Authorization
  LRAD w/Authorization
	Non-Lethal Weapon

Taser	Taser
Canine	Canine
Baton/Expandable baton Impact munitions	Baton
Revolver Rifle Semi-auto pistol Shotgun	Firearm
Other	Other Use of Force

Source: Chicago Police Department TRR Form

Figure A1: Chicago Police Districts (Geographic Units)
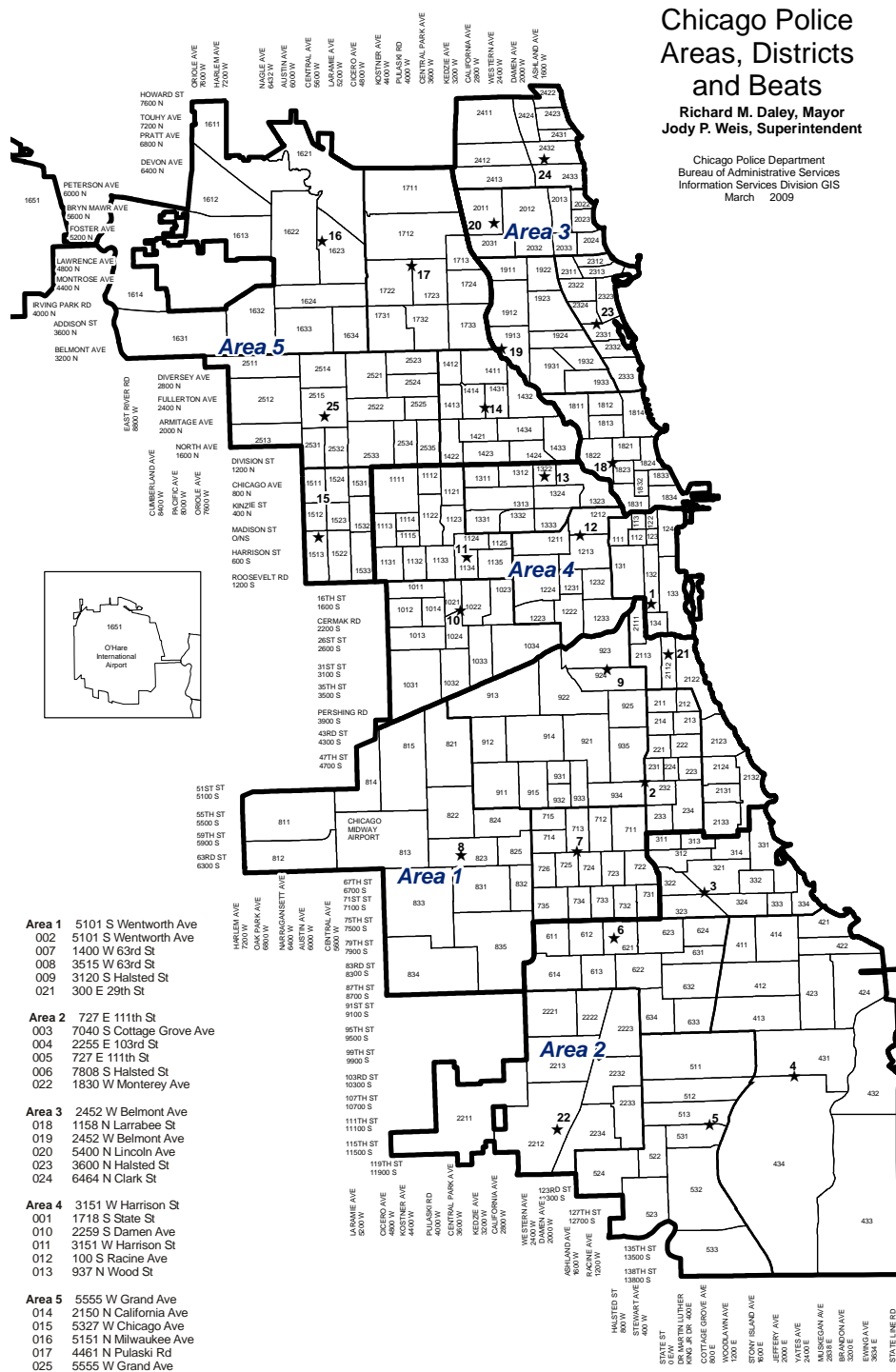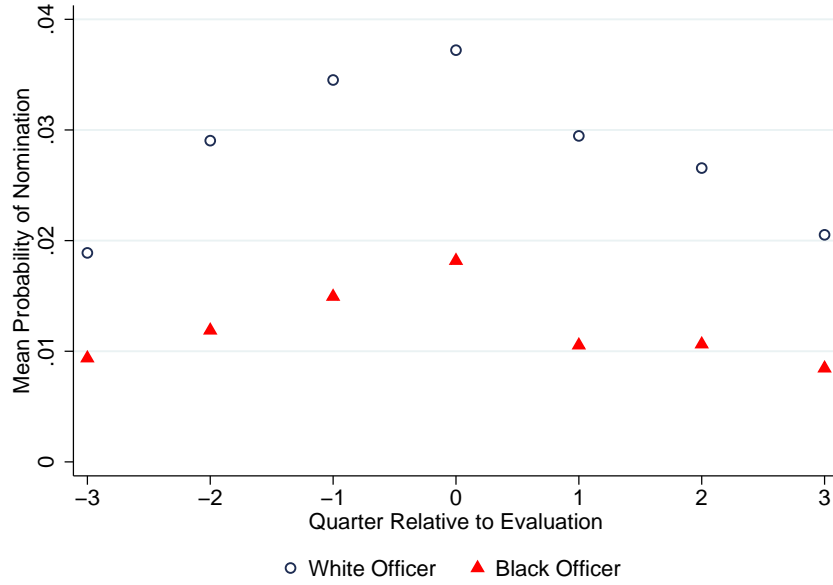

A-3

Figure A2: Average Probability of Nomination by Officer Race and Quarter

Notes: This figure graphs the mean probability of nomination by quarter separately for white officers and black officers. The sample is restricted to officers assigned to white supervisors.

Table A3: Relative Evaluation Quarter

| | Evaluation Quarter | | | |
Month	Q1	Q2	Q3	Q4
1				
2	$RQ = 0$	-1	-2	-3
3				
4				
5	1	0	-1	-2
6				
7				
8	2	1	0	-1
9				
10				
11	3	2	1	0
12				

Note: This table illustrates the relative quarter by month and evaluation quarter. For example, an officer whose evaluation quarter is in Q1, will have January through March classified as $RQ = 0$, April through June classified as $RQ = 1$, July through September classified as $RQ = 2$, and October through December as $RQ = 3$.