The Effect of Minority Peers on Future Arrest Quantity and Quality

Roman Rivera*

Columbia University

Version Date:
June 02, 2022

Abstract

Increasing minority representation in police departments is a common proposal. This paper documents the effect of minority peers in the Chicago police academy on officers' future arrests by exploiting the lottery system which provides exogenous variation in the composition of academy cohorts. I find that minority (e.g., Black, Hispanic) peers reduce officers' future propensity to arrest minority civilians through a reduction in low-quality, low-level arrests. Additional results suggest that other peer characteristics, such as age and gender, modify the effect of minority peers, which is consistent with peers' preferences for aggressive policing playing an important role beyond their minority status.

*I am grateful to Sandra Black, Bentley MacLeod, and Simon Lee for guidance and advice. For feedback and comments, I thank Amani Abou Harb, Douglas Almond, Bocar Ba, Pat Bayer, Michael Best, Felipe Goncalves, Sakshi Gupta, Michelle Jiang, Jenny Jiao, Dean Knox, Taeho Kim, Bob LaLonde, Claire Montialoux, Jonathan Mummolo, Brendan O'Flaherty, José Luis Montiel Olea, Nayoung Rim, Rajiv Sethi, Miguel Urquiola, and Emily Weisburst, as well as the colloquium participants at Columbia University and the Texas Economics of Crime Workshop. I would like to thank Mohammad Abou Harb, Emma Herman, the Invisible Institute, and Sam Stecklow for their contributions to this data set and Rachel Ryley for sharing assignment data. For detailed explanation of the Cook County court system, I thank Ali Ammoura. Email: r.g.rivera@columbia.edu

1 Introduction

Aggressive policing, such as the excessive use of force and over-policing of low-level crimes, has numerous social and economic costs and disproportionately affects minority communities. Increasing minority representation in policing is one of the most common policy proposals to address aggressive policing, as well as to build community trust and to improve police legitimacy (DOJ (2016)). Such policies address the disproportionately white and male composition of police departments relative to the communities they police, as highlighted in Keller (2015) who finds a 24 percentage point gap in minority representation. Existing research finds that minority officers police minority civilians less aggressively (Hoekstra and Sloan (2020), Goncalves and Mello (2021), Ba et al. (2021)). However, the effectiveness of these policies depends on both the direct effect of these officers as well as the spillover effects of increased diversity. Minority officers may influence their peers' policing in myriad ways: interracial friendships may reduce racial bias; negative interactions can lead to animus and prejudice; or officer's may adopt their peers' preferences for policing behavior, such as aggressive policing. Thus, while the full effect of increasing diversity in policing hinges on which of these mechanisms is at play, we still know very little about how peer diversity affects officer behavior.

In this paper, I provide novel evidence of increased shares of minority officers reducing their peers' aggressive policing and improving their peers' arrest quality. This indicates that increased diversity has positive spillovers in policing. Such evidence has previously proved elusive for two main reasons. First, it requires peer groups to be quasi-randomly assigned to avoid self-selection. Second, it requires data on individual police officer demographics, arrests and arrest quality, and peer groups, that are difficult to obtain. I overcome both of these obstacles using detailed data on police officers in the Chicago Police Department (CPD) who were randomly assigned to peer groups in the form of police academy cohorts based on lottery numbers. The police academy is a highly relevant environment: increased departmental diversity requires training increasingly diverse cohorts; the academy forms recruits' first major experiences and relationships in policing; and the CPD academy involves training with one's cohort for 6 months (900 hours), making the peer diversity treatment significantly more intense than most diversity focused interventions.

I first document that officers assigned to police academy cohorts with higher shares of minorities (Black, Hispanic, Asian, and Native American) make fewer arrests of minorities (which represent $> 90\%$ of arrests in Chicago) once they become full officers. However, I find that cohort composition also has a minor influence on where officers work post-academy, which influences arrest opportunities. To correct for this, I control extensively for officers' working conditions and recover individual officer propensities to make arrests of various types using a novel data set on millions of daily shifts and assignments. Consistent with the previous result, higher shares of minorities in academy cohorts reduce officers' future propensities to arrest minorities, especially Black civilians, even after taking working environment into account.[1] Notably, I do not only study the effect of minority peers on whites, as is common in the peer diversity literature. Rather, I study the effect of minority peers on all officers to understand the full extent of spillovers, as minorities make up sizable shares of police departments.

A decline in arrests may be due to either decreased public safety or a reduction in potentially harmful and overly aggressive policing. To distinguish between these two possibilities, I dissaggregate arrests into arrests for serious (e.g., violent, property) and low-level (e.g., drug, traffic) crimes. As serious arrests are crucial for maintaining public safety, a reduction is generally undesirable. However, there is mounting evidence that arrests and prosecution of low-level crimes can significantly harm individuals and actually lead to future criminality (Agan, Doleac, and Harvey (2021)).[2] I find that minority peers have a large negative effect on an officer's propensity to make low-level arrests while having a small positive or null effect on serious arrests. For example, a 5pp increase in Black peers ($\approx$ 1 SD) decreases an officer's propensity to arrest Blacks for low-level crimes by 0.16 standard deviations. Furthermore, this effect is enduring as it is present even four years after the academy has ended.

Beyond the type of an arrest, measuring the 'quality' of an arrest is important as well. If minority peers cause large reductions in productive arrests, then we can interpret the effect as a reduction in the quality or effort of officers. To measure arrest quality, I link arrest records to court outcomes and recover officers' propensities for high (guilty finding in court) and low (no guilty finding) quality arrests for serious and low-level crimes. Prior work measuring arrest quality is limited due to the difficulty in obtaining and linking detailed arrest and court records.[3] I find that the decline in propensities to arrest Blacks for low-level crimes is driven almost entirely by a decline in low quality (not found guilty) arrests with little to no effect on high quality (found guilty) arrests. Combined with the small and positive effects

[1] It is unsurprising that arrests of Black civilians are driving the results, as they make up the vast majority of new officer arrests ($> 80\%$) and differences in enforcement activity between officers tend to be most salient in enforcement against Blacks even among non-Black officers (e.g., male and female, white and Hispanic) (Ba et al. (2021)). However, minority peers have a negative effect on low-level arrests of all groups (e.g., white civilians), though not all point estimates are precisely estimated.

[2] See also Aizer and Doyle (2015), Gupta, Hansman, and Frenchman (2016), Stevenson (2018), and Dobbie, Goldin, and Yang (2018).

[3] Weisburst (2020) also uses court outcomes as a metric for arrest quality. However, my data allow me to estimate an individual officer's propensity to make high and low quality arrests within race and crime-type. Ater, Givati, and Rigbi (2014) also discusses arrest quality, but only use whether the arrest led to a charge as a measure of quality.

on serious arrests, this indicates that minority peers reduce aggressive policing (low-quality low-level arrests) and increase average arrest quality, while not negatively affecting officer effort towards serious crime.

Understanding the mechanisms driving these effects is crucial to develop appropriate policy. As mentioned above, there are multiple potential effects, and I find evidence consistent with two. First, I find that minority peers cause the largest decreases in aggressive policing in white officers. This is consistent with positive interracial contact between whites and minorities reducing bias against Blacks, leading to less aggressive policing. As expected, Black peers have the largest effect on whites with respect to low-level arrests of Black civilians, but other minority peers (Hispanics and other non-whites) also reduce whites' aggressive policing of Blacks. This latter effect may be a result of non-Black minorities facilitating Black-white socialization or positive contact between whites and other minorities indirectly reducing anti-Black bias.[4]

Second, while peer diversity studies often conceptualize peer effects as operating in one direction by examining the effect of minorities (the treatment) on whites (the unit of study), there is evidence that all peers influence each other's actions through forming group cultures, influencing social identity, and altering each others' preferences and personalities (Akerlof and Kranton (2000), Fryer and Torelli (2010), Anwar, Bayer, and Hjalmarsson (2019), Golsteyn, Non, and Zölitz (2021)). I find evidence consistent with this. All officers, white and minority, are influenced more strongly by peers whose observable characteristics are associated with less aggressive policing. This is consistent with officers assigned to cohorts with peers who have lower propensities for aggressive policing also policing less aggressively in the future. Based on the magnitudes, the results suggest that the effect on all officers drives the results and is larger than the whites-only effect.

This paper builds on the literature on diversity and policing. Prior research finds that minority and female representation influences city-level policing outcomes (McCrary (2007), Miller and Segal (2018)), and individual officer race and gender are associated with differential policing behavior (Goncalves and Mello (2021), Ba et al. (2021)).[5] This paper extends this research by identifying peer diversity as a causal determinant of officer's future policing

[4]Positive contacts with one minority group can cause spillovers of improved sentiment toward other minority groups through the secondary transfer effect (Pettigrew and Tropp (2006), Pettigrew (2009), Tausch et al. (2010)).

[5]Notably, McCrary (2007) no effect of affirmative action litigation on crime, but does find a decrease in Black arrest share among serious arrests, where as this paper documents larger peer effects on low-level, not serious, Black arrests. For the effect of representation on more macro-level outcomes see Donohue III and Levitt (2001), Garner, Harvey, and Johnson (2019), and Harvey and Mattia (2019), and Cox, Cunningham, and Ortega (2021). For the more on the relationship between officer race and policing outcomes see West (2018) and Hoekstra and Sloan (2020).

behavior. One implication is that the changes in department-level outcomes resulting from diversity initiatives were the result of both the direct effect of more diverse officers and the indirect effect on their peers. I also build on this literature by studying how peer diversity affects officers' arrest quality, a new outcome in this literature, in addition to their arrest quantity.

This paper also contributes to the growing literature on officer-level interventions. For example, Owens et al. (2018) find that officers assigned to meetings with supervisors were 12% less likely to make an arrest. By comparison, I find a similar reduction can be achieved for low-level Black arrests through a 6.2 percentage point increase in older minority peers in the police academy.[6]

By providing evidence for two main mechanisms, I also contribute to the literatures on both interracial peer effects and the effects of social identity on behavior.[7] First, a substantial literature in psychology and economics studies how minorities, often Blacks, influence the perceptions, biases, and beliefs of whites. They generally find that increased interracial socialization with minorities improves whites' perceptions of minorities (Boisjoly et al. (2006)), increases openness to future contact with minorities (Carrell, Hoekstra, and West (2019)), reduces anti-minority decisions (Anwar, Bayer, and Hjalmarsson (2012)), and reduces implicit bias and decreases participation in racist politics across generations (Schindler and Westcott (2021)).[8] In line with this, I find that minority peers cause larger decreases in aggressive policing of Black civilians among white officers relative to minority officers.

Second, peer diversity studies generally focus solely on the effect of minorities on whites. I expand on this by studying the peer effect of diversity on minorities as well. A smaller literature studies how peers influence outcomes through shifts in preferences and social identity.[9] For example, Anwar, Bayer, and Hjalmarsson (2019) find that jurors' political alignment influences trial outcomes through changing peer opinions. I provide multiple findings consistent with the effect of peer preferences on officer behavior, most centrally that minority peers influence both white and minority officers.

[6] Older being defined as > 32 years, and relative to the share of young (< 27 year) white recruits.

[7] I employ a common identification strategy (random assignment of students to classrooms) in the educational peer effects literature (Sacerdote (2011)) but in a new setting. See Hoxby (2000) and Sacerdote (2001). Angrist (2014) discusses various studies in the educational peer effects literature. Holz, Rivera, and Ba (2019) studies police academy cohorts as well, but their identification hinges on a difference-in-differences design, similar to Ager et al. (2021) which studies pilots. Gould, Lavy, and Paserman (2009), Lavy and Schlosser (2011), Black, Devereux, and Salvanes (2013), Carrell, Hoekstra, and Kuka (2018), and Brenøe and Zölitz (2020) provide evidence for peer composition in educational environments influencing future outcomes.

[8] See also Pettigrew (1998), Laar et al. (2005), Sommers (2006), Pettigrew and Tropp (2006), Baker, Mayer, and Puller (2011), Burns, Corno, and Ferrara (2015), Merlino, Steinhardt, and Wren-Lewis (2019), and Billings, Chyn, and Haggag (2021).

[9] See Akerlof and Kranton (2000), Austen-Smith and Fryer (2005), Akerlof and Kranton (2010), Benjamin, Choi, and Strickland (2010), Golsteyn, Non, and Zölitz (2021), and Holden, Keane, and Lilley (2021).

This paper proceeds as follows. In Section 2, I describe the background and data for this paper. In Section 3, I discuss the empirical strategy, and Section 4 presents the results. Section 5 contains robustness checks, and Section 6 explores mechanisms and alternative explanations. Section 7 concludes.

2 Background and Data

2.1 Chicago Police Department and Recruitment

2.1.1 Application to CPD and the Academy

Comprised of over 10,000 officers, the Chicago Police Department (CPD) is the second largest police force in the US. It polices the nation's third largest city, which is racially diverse and economically segregated. To recruit new officers, the CPD issues a call for officers, and applicants take a written exam, which they must pass in order to enter the academy. As a CPD Frequently Asked Question (FAQ) form, CPD (2017), explains:

> All applicants who pass the exam are placed on an eligibility list based on a randomly assigned lottery number. You will be referred to the Chicago Police Department in lottery order as vacancies become available.

After an applicant's number is called, if they pass required physical and psychological tests, they are permitted to start at the police academy (see Appendix A.1 for more discussion). Academy start dates, known as "appointed dates", correspond to officers beginning their time at the police academy. In Appendix A.2, I provide empirical support for the random assignment of officers to cohorts. I define a cohort as the group of officers with the same appointed date—in the main sample, cohorts are separated by about 1 month. During the academy, officers must complete 900 hours (about 6 months) of training in multiple areas, such as "firearms, control tactics, physical training, [and] classroom training" (CPD (2020)).[10]

After the academy, the recruits in a cohort enter an on-the-job-training period for one year as "probationary police officers" during which they are split up, work in multiple areas of the city, and are evaluated under the supervision of Field Training Officers. After recruits meet the various requirements, complete their time as a probationary officers, and become "field qualified" (CPD (2018)), they exit their probationary period and become a full (sworn)

[10]The 900 hours of training encompasses and surpasses the training required to pass the Illinois State Peace Officer's Certification Exam. In larger cohorts, officers are further subdivided into "homerooms" which take most of their trainings together– while data on homerooms could not be obtained, I use detailed training data to approximate these groups.

Chicago police officer. New sworn officers are then assigned to more permanent units.

2.1.2 Sample and Requirements

This paper focuses on the cohorts with start dates between 2009 and 2016, as they overlap best with the data. These cohorts can be divided into three periods based on the year during which officers took the entrance exam, i.e. the level at which they were randomly assigned lottery numbers. (See Figure B.1 for exam information.) In 2006, three exams were given in rapid succession, each attended by a relatively small number of applicants (all between 800 and 1,500 passing applicants)– these tests will be collectively referred to as Exam 2006. The next test was issued in 2010 (Exam 2010), with almost 8,000 passing applicants. In 2013, the final test in our sample was issued (Exam 2013) with over 12,000 passing applicants and an important policy change: the minimum age of entry was reduced from 23 to 21 (Pritchard (2013))– the maximum age is 40 years old for all cohorts in the sample. While the CPD did not provide information on which officers belong to which test, in Appendix A.1 I show that the Exam 2010 cohorts can be identified with near certainty.

The CPD is massively over-subscribed: fewer than 3,000 applicants were called into the academy between 2009 and 2016, while over 20,000 applicants passed the exams. This is because CPD jobs are highly desirable by individuals from across the country, and applicants to law enforcement are highly passionate about joining a police force. In this sample, the CPD began to call individuals into the academy years after their respective exam was taken: the last batch of 2006 test-takers were likely called between March 2009 and October 2011; 2010 test-takers were first called into the academy between April 2012 and May 2014; and 2013 test takes were likely called between August 2014 and December 2016.[11]

2.1.3 Units and Daily Assignments

Transfers between assignments and the filling of vacancies are determined by a seniority-based bidding process and are only available to non-probationary sworn officers, meaning new officers have little to no choice in where and when they work (CPD (2011)). New officers are generally assigned to the patrol units which correspond to geographical districts in Chicago.[12] These units occupy most CPD officers and relate to what is commonly considered police work. There are many other units for specialized work that contain far fewer and more

[11]Based on internet discussions, passing applicants with a high lottery number (far into the queue) are advised that "They will probably test again before they call you" and that "It may take another year or so but you will keep moving up the list... Just keep training and stay strong and good luck!!!" (feredeathpsn (2017)).

[12]There were 25 districts / patrol units before 2012. During 2012, three of these units/districts were collapsed into other districts, reducing the total number to 22.

experienced officers, such as training units, detective units, etc., which are not studied in this paper.

Within units, officers also bid for shifts/watches (generally, 12 am - 8 am, 8 am - 4 pm, 4pm - 12 am), their 'day off group', and furlough days at the end of the preceding year– this is also seniority based. On any specific day, whether or not an officer is assigned to work depends on their rotating schedule, which is generally 4 days on and 2 days off during the period of study, and is predetermined by their day off group based on the CPD operations calendar (see Figure B.2 and Ba et al. (2021) for more details). This means that the exact days an officer works are predetermined, not up to the officer's discretion on that day, and rotate over the days of the week.[13] Furthermore, it means that the exact composition of the officers working in a unit and watch on a specific day will be different the following week as officers of different day of groups and furlough schedules will be working together– officers do not frequently work in different shifts throughout a year.

2.2 Data

The data for this study come from the Chicago Police Department, Chicago's Department of Human Resources, and the Circuit Court of Cook County (in which Chicago is located). By combining data sets on CPD officers obtained over five years, I construct a detailed panel data set of officer assignments, arrests, and arrest outcomes in court between 2010 and 2018. This contains officers' demographic information (race, gender, age), start dates, when officers exited the training unit (after the academy and probationary period), and other administrative information. Daily assignment and attendance data includes daily records on officer assignments and time on duty for the geographic units. Additional data sets contain information on trainings, officer education, military status, and language ability. Collectively, these data permit highly granular analysis of an officer's working environment and peer groups. I restrict my analysis to observations of police officers (the lowest and most common rank, e.g., not detectives, sergeants, etc.) working on shift (watch) numbers 1-4, and assigned to regular assignments (e.g., not administrative, lockup, desk duty, etc.).

In order to recover individual officer's arrest quantity and quality metrics, I use arrest data and court data. The arrest data contain all arrests of adults by Chicago police officers including arrest date and time, crime description, primary arresting officer(s), and arrestee race.[14] By linking the arrest data to court records, I construct a metric for arrest quality by

[13] For example, one week an officer works Tuesday through Friday, and the next week they work Monday through Thursday

[14] Almost all arrests have at most two primary officers listed. Data on juvenile arrests is much more limited as it is protected from FOIA; for example, central booking number, race, gender, and crime type are

determining if the arrest was associated with any guilty finding, which indicates high quality, or no guilty, which indicates low quality. Guilty findings include plea deals, which account for over 90% of convictions. Combined, these data allow me to construct a measure for individual officer's arrest quantity and quality after extensively controlling for their working environments. For more discussion of the data, see Appendix C.

2.2.1 Sample Selection

A total of 2,795 officers joined the CPD between March 2009 and December 2016. As defined above, an academy cohort is all the recruits who started at the CPD academy on the same date, resulting in 69 cohorts during this period. I focus my primary analyses on the Exam 2010 cohorts (the "Main Sample"), starting between July 2012 and May 2014, for multiple reasons. First, I can observe main sample officer assignments and arrests from their probationary periods onward, and during their time at the academy (July of 2012 to mid-2015), there were no changes in departmental leadership or major political or policing scandals in Chicago. Relative to the other exams, the assignment patterns of Exam 2010 cohorts are most consistent the random assignment assumption (see Appendix A.2), and main sample cohorts originated from the same entrance exam issued in December of 2010 (see Appendix A.1), which is not as certain for the 2006 test or 2013 test cohorts. Furthermore, there are twice as many officers in Exam 2010 relative to Exam 2006, and Exam 2010 officers can be observed twice as long into their careers relative to Exam 2013 officers. Lastly, I am able to include data on field training officers for Exam 2010, but not the other cohorts. The downsides to using only the main sample are that there will be fewer cohorts and that cohorts will be the only level of variation. I refer to Exams 2006, 2010, and 2013 collectively as the "Full Sample".

All officers in the full sample were subject to a series of filters.[15] Notably, I drop recruits in cohorts who were not matched in the assignment data, recruits with invalid durations in the academy or probationary period, and recruits not matched in the salary and unit assignment data. I also drop a few recruits that had fewer than 15 observations in the assignment panel. Attrition from the initial cohort to the final sample can occur for multiple reasons. If attrition is related to cohort composition, it may contaminate the results. But, as I show in Appendix A.3, cohort diversity has no significant impact on attrition in the

redacted.

[15]Before calculating initial cohort composition, I excluded very small cohorts that started during the sample period but had cohorts with fewer than 7 recruits which removed a total of 37– the majority were in single officer cohorts. These small 'cohorts' are likely errors as the next smallest cohort size is 25. I also dropped 1 recruit who reported starting too young, and another 9 officers who likely had erroneous start dates were also removed.

main or full samples. After filters, the main sample of cohorts contain 940 new officers in 21 cohorts with 531,597 total officer-shift observations over 61 months. The full sample, likewise, contains 2,336 officers in 43 cohorts with 1,081,543 total officer-shift observations over 100 months.

2.3 Summary Statistics

2.3.1 Cohort Composition

Table 1 displays the demographic composition of each exam period (Exam 2010, 2006, 2013) in Columns (1)-(6) with even columns containing pooled means and odd columns containing means over cohort compositions before attrition. Column (7) contains the pooled demographics of all officers in the panel data for reference. By comparing Columns (1) and (2), it is apparent that the main sample of recruits is very similar to that of their average cohort, which is expected due to the random assignment of recruits to cohorts, and that attrition after entering the academy did not significantly alter the demographic composition of the pool of officers. This similarity is also apparent by comparing the pooled and cohort means for the other exam periods. Overall, the main sample of recruits is 80.74% male, start at 30.1 years old, and 48.72% minority– which is comprised of mostly Hispanics (31.17%) and Blacks (13.19%). The average main sample cohort contains 54 recruits.

The comparison between pooled demographics of all officers (Column (7)) and the recruit demographics for each Exam period illustrates the changing nature of the Chicago Police Department. More recent recruits are less likely to be female than all panel officers (24.43% vs. the 2013 cohorts at 23.32%). While minorities make up roughly half of both groups, the composition of minorities has changed: Black officers make up about 22.75% of all panel officers while their share has been decreased by almost half from Exam 2006 (22.42%) to Exam 2010 (13.19%), to Exam 2013 (12.56%). The sharp decline in Black recruitment has been made up for by a surge in Hispanic recruitment (26.1% for all officers vs. 34.08% for Exam 2013). This pattern is generally representative of police departments across the country in the last 30 years (Keller (2015)). For a visualization of cohort compositions see Figure B.3.

Between Exam 2010 and Exam 2013, the reduction in start age requirement from 23 to 21 was associated with about a 2 year decline the average start age of recruits (Columns (1) and (5)). Figure B.4 displays the cumulative CDF of officer start ages in the three exam periods. Notably, in the Exam 2010 and 2006 cohorts, only 27% of recruits started before they were 27, while in the Exam 2013 cohorts 49% did.

The top panel of Table 2 displays summary statistics of main sample cohort compositions for all officers in Column (1), and Columns (2) and (3) divide these officers by whether their cohort had high ($\geq$ 50%) or low ($<$ 50%) minority share (50% minority is the median). Low-minority cohorts are on average 43.8% minority compared with high-minority cohorts at 54.51%, with similar standard deviations– over all cohorts, one standard deviation of cohort share minority is 0.06. Comparisons of other demographic compositions show little differences in observables between low and high minority cohorts. The differences in cohort share female and mean start age are neither statistically significant nor economically large.

2.3.2 Policing Outcomes

Arrests are a common metric when studying individual officer and departmental performance and, in the light of concerns about over-policing — excessive and detrimental interactions between law enforcement and civilians — arrests are the main metric I use to measure officer enforcement activity.[16] To distinguish between the seriousness of arrests, I divide them based on crime: serious arrests, which I define as arrests for official FBI index crimes, as well as additional forms of homicide, fraud, domestic violence, sexual assault, and simple assault and battery; and low-level crimes are all other arrests—e.g., warrant, traffic, or drug crimes.[17] I also classify arrests based on arrestee race/ethnicity (white, Black, Hispanic, or other). Using Cook County court data, I determine if the arrest is associated with a guilty finding, and I interpret this as a measure of arrest quality.[18] Opportunities for officers to make arrests depend on the crime rates where and when they work, which influence the

[16]Arrest counts (and the clearance of crimes) are a common metric of police activity. See Donohue III and Levitt (2001), Mas (2006), McCrary (2007), Shi (2009), Coviello and Persico (2015), Blanes i Vidal and Kirchmaier (2018), Owens et al. (2018), Garner, Harvey, and Johnson (2019), Weisburst (2020), and Kirchmaier et al. (2021).

[17]Index crimes are offenses on which the FBI collects data and tracks and publishes annually in the Uniform Crime Report (UCR). The eight index crimes are four violent and four property offenses: (violent) aggravated assault, robbery, murder, rape, (property) burglary, larceny, motor vehicle theft, arson. For non-index crimes I classify as 'serious', domestic violence is determined by whether the description indicates domestic battery or assault, and a few additional sexual assaults were classified based on whether the description indicates criminal sexual assault. Simple assaults and battery include crimes such as attempts at assault, child abuse, and threats of violence. I classify multiple types of deceptive practices as fraud. As a robustness check in Section 5, I redo the main analysis using the FBI index and non-index crime distinctions for serious and low-level crimes, respectively, and find similar results.

[18]I define an arrest to be 'guilty' if the central booking number (CBN) is associated with any guilty finding; I consider an arrest not guilty if the CBN is associated with no guilty findings and at least one not guilty finding. If a CBN is associated with no guilty findings and no not guilty findings, and it has any dismissed cases, then I consider it dismissed. If a CBN does not appear in the court data, I classify the case as dropped. I group not guilty, dismissed, and dropped cases together and label them as 'non-guilty'. If a CBN is not classified as guilty, not guilty, or dismissed, but it is in the court data, then it only has incomplete/open cases, so it is classified as neither guilty nor non-guilty. A single CBN may have multiple charges or cases associated with it, and I use the method discussed above to provide a single outcome of an arrest which is conservative as only one guilty verdict on any charge is sufficient for an arrest to be 'guilty'.

quantity, quality, and kind of arrests.[19]

The bottom panel of Table 2 displays arrests per shift, violent index crime rates, and observations in the daily panel data for all main sample recruits as full officers in Column (1), and Columns (2) and (3) divide these officers by whether their cohort had high ($\geq 50\%$) or low ($< 50\%$) minority share.[20] Note that all differences in arrests and violent crime between Columns (2) and (3) are statistically significant. The vast majority of arrests are of Black civilians (81.8%), with Hispanic arrests being far less common at 13.2%. Recruits in high-minority cohorts make fewer arrests per shift than those in low-minority cohorts, driven by a difference in arrests of Blacks (0.1378 vs. 0.1481). About 70.31% of arrests are for low-level crimes. Recruits in low-minority cohorts make slightly more guilty arrests relative to those in high-minority cohorts, with 30.64% and 28.9% guilty, respectively. Recruits in low-minority cohorts work, on average, in slightly lower crime districts relative to recruits in high-minority cohort, yet both groups work in Chicago's most dangerous areas.[21] While this table documents differences between new officers in terms of arrest quantity and quality, as well as working environment based on cohort diversity, whether cohort diversity is actually changing officer enforcement behavior requires more detailed analysis.

3 Empirical Strategy

3.1 Peer Effects Framework

The aim of this paper is to estimate the long-run effect of peer diversity on officer behavior. The identification strategy for this paper borrows heavily from the education literature on long-run peer effects, leveraging the random assignment of officers (students) to academy cohorts (classrooms). As a first step, I adapt the regression specification from the long-run peer effects in education literature (Chetty et al. (2011), Carrell, Hoekstra, and Kuka (2018)) by regressing outcomes on the characteristics of randomly assigned peers. Specifically, I estimate:

$$\overline{Arrest}_i^k = \alpha_{p(i)}^k + \pi_1^k \overline{X}_{c(i)} + \pi_2^k X_i + v_i^k \tag{1}$$

where $\overline{Arrest}_i^k$ is the average number of arrests type k (e.g., Black low-level guilty arrests) per shift made by officer i randomly assigned to cohort $c(i)$ in exam period $p(i)$. Variable

[19]For example, lower crime may mean the marginal arrest is less likely to be high quality if officers value making arrests.

[20]Index violent crimes are murder, rape, robbery, and aggravated assault.

[21]A monthly violent crime rate of about 15 per 10,000 population is the 75th percentile of monthly violent index crime rates in Chicago.

$\alpha^k_{p(i)}$ is a fixed effect for the exam $p(i)$ which officer i took (as did all other officers in i's cohort $c(i)$). X_i contains the demographic characteristics (e.g., race, start age) for officer i. $\overline{X}_{c(i)} = \frac{\sum_{j \neq i}^{n_c} X_j}{n_c - 1}$, contains the leave-out mean of the demographic characteristics of members of officer i's cohort c.[22]

The random assignment of lottery numbers within an exam pool allows cohort composition, $\overline{X}_{c(i)}$ to be uncorrelated with unobserved characteristics about the officer, v^k_i, permitting consistent estimation of the peer effect of cohort diversity, π^k_1 (see Appendix A.2 for tests of random assignment). More formally: $\mathbb{E}[v^k_i | \overline{X}_{c(i)}, \alpha^k_{p(i)}] = 0 \; \forall \; i$.[23] However, the mechanism by which cohort composition influences future arrests is not specified. One part of π^k_1 is the effect of cohort diversity on an individual officer's behavior, opinions, beliefs, and prejudices. Yet, as cohort diversity influences officer assignments and future peers (discussed more in Appendix A.4 and Appendix A.5), and assignments influence arrest possibilities, the other part of π^k_1 is the assignment effect of diversity. So, though the assignment effect proves to be minor, π^k_1 is a causal estimate of the effect of cohort composition on an officer's future arrests of type k *within* the assignment system of the Chicago Police Department.[24]

3.2 Officer Heterogeneity

Using a raw arrest metric, e.g., arrests per shift, as the outcome of interest is straightforward, but it suffers multiple issues. First, higher shares of minorities in cohorts cause officers to make fewer arrests of Blacks during their careers. This is in part due to how cohort diversity influences new officer assignments and how officers choose to bid for assignments (see Appendix A.4). Second, different cohorts do not start at the academy at the same time, and their opportunities for arrests will be influenced by departmental demand, non-linear changes in crime rates, and other factors, making arrests alone a highly noisy outcome.

For the effect of peers to be externally valid and relevant for police departments with different priorities and assignment policies, understanding how peers influence officer's individual *type* is necessary. By officer type, I mean their individual propensity to make an arrest, a measure of their enforcement activity regardless of their working environment, or

[22]For computing $\overline{X}_{c(i)}$, I include all recruits beginning in the cohort c excluding i.

[23]Given that cohort composition is randomly determined and $\overline{X}_{c(i)}$ excludes the officer i, cohort composition excluding officer i is independent of officer i's observable characteristics, X_i. So, leaving out X_i should not impact estimates of π^k_1.

[24]In this setting, I cannot distinguish between endogenous and exogenous effects of peers (Manski (1993)), which means I cannot disentangle the effect of officers being affected by minority peers due to their behavior or their characteristics. I assume there are no correlated effects (e.g., instructor effects)– given the large amount of courses recruits are taught during the academy, it is highly unlikely that a cohort with 40% minority composition would receive different institutional environments or instructors than a cohort with 50% minority composition starting a month later.

their individual contribution to the quantity or quality of arrests they make. Controlling for working environment also alleviates concerns about differential assignments and crime rates over time.

Identifying the effect of peers on officer types follows two steps. First, I recover a measure of an officer's 'type', i.e. their propensity to make arrests net of high-dimensional daily assignment fixed effects and other factors. Second, I regress these arrest propensities on cohort composition to estimate the long-run effect of peer diversity on individual officer behavior. This allows for data reduction and exploration of heterogeneity, and it permits flexible specifications in the first stage.[25]

I first recover an estimate for all officers' (including those outside the full sample of cohort officers) propensities to make arrest of type k, θ_i^k, using a first stage regression.[26] I estimate a linear fixed effect regression model:

$$Arrest_{it}^k = \theta_i^k + \gamma_{brsw_t}^k + \beta^k V_{it} + \epsilon_{it}^k \tag{2}$$

where $Arrest_{it}^k$ is the number of arrests of type k officer i made during their on-duty time on date t.[27] The data is sufficiently rich such that I can control for a large set of assignment and environment characteristics with highly specific fixed effects, $\gamma_{brsw_t}^k$, which interacts officer i's assigned district and truncated beat code (b), their role (r), their shift number (s), and the year, month, and day of the week (w_t).[28] The data has over 7.8 million officer-shift observations on about 14,000 officers and contains approximately 580,000 assignment fixed effects (γ_{brsw_t}).[29] V_{it} controls for second-degree polynomials of officer i's tenure. All random

[25] Weisburst (2020) uses an analogous method by first recovering officer fixed effects for making arrests following 911 calls, then regressing these officer fixed effects on officer characteristics, and Card and Krueger (1992) use a similar two-step procedure for studying the determinants of returns to education.

[26] More formally, $k \in \begin{pmatrix} All \\ Serious \\ low\text{-}level \end{pmatrix} \times \begin{pmatrix} All \\ Guilty \\ Non\ Guilty \end{pmatrix} \times \begin{pmatrix} Minority \\ Black \\ NonBlack \\ Hispanic \\ White \end{pmatrix}$.

[27] Estimation was performed using the R package 'lfe' (Gaure (2013a)), which implements the algorithm introduced in Gaure (2013b) that is designed for estimating linear models with multiple overlapping high-dimensional fixed effects (e.g. officers and who move across shifts or workers who move across firms). Notably, this package also allows for standard errors of the fixed effects to be recovered which is used in Appendix A.6.

[28] Formally, b is the numeric beat code with the last numerical digit removed, and r is the exact role designated by the full beat code. For example, beat code "2533" has a role of 'beat officer' and the beat is truncated to "253" which indicates the sector they work in (a group of contiguous geographic beats, and beats are on average less than 1 square mile). Beat code "2463A" has a role of tactical team C officer, as does beat code "2463C", and both have the same truncated beat as "246" (which does not map to a geographic sector), so their $brsw_t$'s are identical if they also work in the same watch, day of week, month, and year.

[29] While the assignment fixed effects are highly granular, there is still significant variation within and across assignments. There are on average 8 officers per value of γ_{brsw_t}, and on average officers work in 9 different b's and 5 different r's, for example. Thus, there is sufficient variation in assignments across and within officers

shocks to an officer's arrest participation during their working period are contained in ϵ_{it}^k.

I assume that conditional on a polynomial of officer tenure and officer assignment fixed effects: 1) current and future shocks to arrest counts are orthogonal to past observables; 2) shocks to arrest counts are not serially correlated across shifts; and 3) since the number of daily shifts I observe for each officer grows quickly, the officer fixed effects are consistently estimated (the mean number of observations for an officer in the panel is 558). I interpret the recovered $\hat{\theta}_i^k$ as an estimate of the individual officer's propensity for enforcement of type k, which I recover for all officers in the daily assignment panel between 2010 and 2018.

With this first stage regression, I control for significant temporal, geographic, demographic, and income variation in where each officer is working, as well as the within-day heterogeneity, officer exposure to different types of civilians and local crime rates, and the effects of officer tenure.[30] Another strength of this design is that I am able to leverage data on all officers, not just those in the sample cohorts. So, I am using all the variation across officer shifts without having to drop observations from officers whose cohorts I do not observe. This means I have sufficient observations within highly granular assignments to use high-dimensional fixed effects (i.e. $\gamma_{brsw_t}^k$) and allow for interactions between assignment characteristics. Recovering a single metric (per arrest type) for each officer also avoids weighting issues as new cohorts have fewer observations in the panel data than older ones.

Using only the fixed effects of officers in my sample cohorts, I replace $\hat{\theta}_i^k$ as the dependent variable in equation (1):

$$\hat{\theta}_i^k = \alpha_{p(i)}^k + \pi_1^k \overline{X}_{c(i)} + \pi_2^k X_i + v_i^k \tag{3}$$

Now, π_1^k can be interpreted as the peer effect on an officer's propensity to make arrests of type k.[31] As before, the composition of one's cohort is independent of one's own preexisting characteristics, but now the outcome is the result of extensively controlling for working environment such that $\hat{\theta}_i^k$ is officer i's individual contribution to make arrests of type k regardless of when or where they work. The minor effect of assignment crowding out due to cohort diversity is removed from this measure, and the exogeneity assumption ($\mathbb{E}[v_i^k|\overline{X}_{c(i)}, \alpha_{cp}^k] = 0 \forall i$) holds, making π_1^k the causal effect of cohort diversity on officer enforcement propensity. $\hat{\theta}_i^k$ is an estimated quantity (and thus has measurement error), and it is standing in for the ideal outcome variable in equation (3), θ_i^k, which is the officer's true but unobserved type. Following the teacher value added literature, I apply a Bayesian

to identify officer fixed effects.

[30] The CPD's operational schedule reinforces the inability of officers to select shifts on specific days or civilian pools (see Ba et al. (2021) for more detail).

[31] I assume a homogeneous treatment effect across officers.

shrinkage procedure to the fixed effects to obtain more precise fixed effects and reduce measurement error (see Appendix A.6).

4 Results

4.1 Main Sample Results

The recovered distributions of main sample officer fixed effects indicate differences across race and exposure to diversity.[32] Figure 1 presents graphical evidence for heterogeneity in officer enforcement being related to officer race and cohort diversity. The distribution of officer arrest propensities (fixed effects) has a long right tail, as in Weisburst (2020). Panel A displays the distributions of fixed effects for arresting Blacks for white and minority officers in the main sample. White officers tend to have higher fixed effects, i.e. a higher individual propensity to arrest Blacks, relative to minority (non-white) officers. This conforms with existing research on white officers policing Blacks more aggressively.[33] Panel B displays the distribution of fixed effects for white officers in the main sample split by cohort share minority. Clearly, white officers in cohorts with more minorities tend to have lower fixed effects—lower individual propensities to arrest Blacks.

4.1.1 Effect on Minority Arrests

For more detailed results, I turn to regression analysis. Table 3 displays the central results for the main sample. First, Column (1) displays the negative relationship between cohort share minority (CSM) and main sample officers' average arrests of minorities in their first 200 shifts (slightly over one year), estimated using equation (1). The coefficient on CSM, -0.188 ($p = 0.057$). This indicates that a 10pp increase in CSM (about 1.6 SDs) is associated with 4 fewer arrests of minorities over their first 200 shifts, equivalent to a 12% decline relative to the mean. Column (2) displays the effect of CSM on the officer's fixed effect (individual propensity after controlling extensively for assignment and temporal effects) to arrest minorities, estimated using equation (3). The coefficient is not statistically significant and smaller in magnitude (-0.138, $p = 0.11$), which is expected because of the negative bias in estimating equation (1) due to high CSM reducing the level of crime and minority population in an officer's working district (as shown in Appendix A.4).

The imprecision of Column (2) is due to significant underlying heterogeneity in the

[32] I solely discuss the fixed effects for officers in the main sample in this section, so "officer" or "recruit" both refer to officers in the main sample as sworn/full officers after their probationary period.

[33] See Goncalves and Mello (2021), Hoekstra and Sloan (2020), Weisburst (2020), and Ba et al. (2021).

effect of minority peers on officers' propensities to arrest minorities. As shown in Columns (3) and (4), CSM has a large negative effect on propensity to arrest minorities for less serious (discretionary) crime but has a small positive effect on propensity to arrest minorities for serious (property and violent, non-discretionary) crimes. Based on Column (3), the coefficient on CSM (-0.186, $p < 0.05$) indicates that a 10pp increase in CSM decreases an officer's propensity to arrest minorities for low-level crimes equivalent to 1.86 fewer arrests over 100 shifts, which corresponds to a decrease 0.24 standard deviations and a 11% decline relative to the mean. Column (4), on the other hand, indicates that a 10pp increase in CSM increases an officer's propensity to arrest minorities for serious crimes by 0.48 more arrests over 100 shifts which is equivalent to 0.19 standard deviations. Columns (5) and (6) replicate Columns (3) and (4) with full controls for officer race, gender, start age, and cohort size; the results are almost identical with coefficients and standard errors increasing slightly. The robustness of the coefficients to these controls is further support for random assignment of officers to cohorts in the main sample.

The coefficients for officer race, gender, and start age, in Column (5) indicates that, for arrests of minorities for low-level crimes, relative to white officers, Black officers make 3 fewer arrests per 100 shifts while Hispanic officers make 1.6 fewer; male officers make 2.2 more arrests than female officers per 100 shifts, and officers who start the academy at one year older make 0.2 fewer arrests per 100 shifts. These differences by officer characteristics decrease in magnitude and statistical significance for serious arrests (Column (6)), as expected given that officers have more discretion over making low-level arrests as opposed to serious ones. These results are also in line with previous research about differences in aggressive policing by officer race, and the results are quite similar to those of Ba et al. (2021).[34] Table B.1 provides more detailed results on the association between demographics and propensities for aggressive policing.

Columns (7) and (8) repeat the analysis in Columns (5) and (6) with CSM broken down into cohort share Black (CSB) and cohort share non-Black minority (CSN), which includes (mainly) Hispanics, Asians/Pacific Islanders, and Native Americans. Black peers have a significantly larger effect relative to non-Black minority peers on low-level arrests of minorities (-0.314, $p < 0.05$, vs. -0.193, $p < 0.05$) and serious arrests of minorities (0.097, $p < 0.01$, vs. 0.043, $p < 0.05$). The effect of CSB is about 2 times larger than the effect of CSN for both types of arrests.

[34]See Close and Mason (2007), West (2018), Goncalves and Mello (2021), Weisburst (2020), and Hoekstra and Sloan (2020).

4.1.2 Disaggregating Minority Arrests

Thus far, we have considered arrests of minorities as a single group which comprised just over 90% of arrests; however, 77.43% of minorities arrested are Black– the rest are almost all Hispanic (21.6%). Columns (9)-(12) break down arrests of minorities into arrests of Blacks ((9)-(10)) and arrests of non-Black minorities ((11)-(12)). Comparing Columns (9)-(10) and (11)-(12) shows that the effects on minority arrests are driven by arrests of Blacks, which maintains statistically and economically significant coefficients. A 10pp increase in CSB (CSN), equivalent to 2.1 (1.3) SDs, decreases officer propensity to make low-level arrests of Blacks by -0.0256, $p < 0.05$ (-0.0153, $p < 0.05$). This is equivalent to a 25.6% (15.3%) decrease relative to the mean. By contrast, the effects on non-Black minority arrests are directionally consistently but are small (for serious) and not statistically significant (for low-level). This is consistent with Table B.1, which shows that race, gender, and age have much stronger influences on Black arrests relative to non-Black arrests.

I focus on Black arrests as the main outcome of interest in the following sections because they drive the minority arrest results and are the vast majority of arrests in Chicago (>80%), though minority peers decrease officer propensities to arrest all civilians for low-level crimes.[35] Furthermore, I shrink officer fixed effects using a Bayesian shrinkage procedure described in Appendix A.6 for more precise results and use shrunken fixed effects in the following analyses.

Columns (13)-(14) replicate Columns (9)-(10), but using shrunken fixed effects. The effect of Black (non-Black minority) peers on low-level arrests shrinks by 22.27% (23.53%) and on serious arrests shrinks by 65% (78.79%), and standard errors decline. The fact that serious arrest fixed effects shrank more substantially than low-level ones is in line with the fact that they are much less frequent.

These final results indicate that a 10pp increase in CSB (CSN) decreases arrests of Blacks for low-level crimes by -1.99, $p < 0.05$ (-1.17,), over 100 shifts, equivalent to a -0.32 (-0.19) SD decrease. For serious arrests of Blacks, a 10pp increase in CSB (CSN) increases arrests of Blacks for serious crimes by 0.28, $p < 0.05$, (0.07, $p > 0.1$ over 100 shifts, equivalent to a 0.18 (0.05) SD increase.

4.1.3 Dynamic Effects

It is important to consider whether these effects are short-term or if they persist into an officer's career. While multiple years of observations are not available for full sample offi-

[35] Table B.2 shows that minority peers' decrease low-level arrests of civilians of all racial groups (Black, non-Black minority (mostly Hispanic), and white civilians), though point estimates are less precise for non-Black groups. The effects are largest and most precisely estimated for Black arrestees, which is consistent with the fact that Black civilians make up the vast majority of arrests.

cers, the main sample officers are observable for a minimum of 3 years into their careers as full officers (4 years after they exit the academy). To explore the dynamic effects of peer composition, I re-estimate equation (2) with individual fixed effects for officers for each 180-day period– meaning an officer during the first 180 days in their career as a full officer is considered a different individual as that same officer in their next 180 day period. Figure B.5 displays the coefficients of the main specification (Column (13) in Table 3) estimated separately for each 180-day fixed effect for main sample officers. The figure indicates that the effect of non-Black minority peers slightly attenuates overtime, while the effect of Black peers remains stable. While 3 years is not an officer's full career by any means, the results indicate that the strongest effect (that of Black peers) is persistent long after the academy classes dissipate.

4.2 Full Sample Results

In this section, I expand the data used to the full sample– including the Exam 2006 and Exam 2013 cohorts. While adding additional observations is useful, the data come with limitations discussed in Section 2.2.1. Table 4 displays the results for the full sample, with the outcomes being shrunken officer propensities to arrest Blacks for low-level crimes (odd columns) and serious crimes (even columns), using equation (3) and including exam fixed effects.

Columns (1)-(2) display the results for Exam 2010 and Exam 2006 cohorts– the results do not change significantly relative to those in Columns (13) and (14) in Table 3. Columns (3)-(4) estimate the effects for the full sample (including Exam 2013). With the inclusion of the full sample, the effects of cohort share Black ($p < 0.01$) and non-Black minorities ($p > 0.1$) both decline and become less precise. Recall that the age requirement policy changed for the Exam 2013 cohorts, with the minimum start age being lowered from 23 to 21. This led to a significant compositional change between Exam 2010 (and Exam 2006) and Exam 2013 cohorts: there was a resulting shift in age distribution and the inclusion of a significantly higher portion of officers below the age of 27, as shown in Figure B.4 and discussed in Section 2.3.1.

Collectively, this raises the question of whether start age or minority status is driving the results. To disentangle the effect of age and minority status, I divide start ages into three groups: young (< 27), mid ($27 - 32$), and old (> 32), then I compute leave-out-means of each age group interacted with minority status (grouping Blacks, Hispanics, Asians, and Native Americans together as in Columns (1)-(6) in Table 3). Then I estimate equation (3) replacing cohort shares with these age and minority interaction groups. Controls include the share of mid and old white recruits, making the reference the share of young white peers. The results are displayed in Column (5)-(8), and they indicate that the interaction between age and minority status is a key factor with respect to the peer effects of diversity.

Based on the results in Column (5), the effect of minority peers is increasing by their age: young minority peers only slightly decrease arrests of Blacks for low-level crimes (-0.061, $p > 0.1$), mid-aged minority peers have a larger but noisy effect (-0.128, $p > 0.1$), and old minorities have an economically and statistically significant effect (-0.189, $p < 0.01$). Column (6) indicates that the effect of minority peers on serious arrests is also driven by older minorities (0.012, $p > 0.1$). This pattern in age groups is consistent with older officers policing Blacks less aggressively (see Table B.1). Columns (7) and (8) replicate (5) and (6) but excluding the Exam 2013 cohorts to ensure the shift in age distribution is not driving the results, and the patterns and results for the effects of minority peers are very similar for low-level arrests.

4.3 Effect on Officer Arrest Quality

A unique feature of my data is that I can observe the outcomes of arrests in court, which enables me to measure the quality of arrests. To study the effect of peer diversity on arrest quality, I classify high-quality arrests as those which result in a guilty outcome in court and low-quality arrests as those which result in a non-guilty outcome in court.[36] Much like arrest quantity, officers are heterogeneous in their arrest quality, and officers that have high propensities to make guilty (high-quality) arrests do not necessarily have high propensities to make non-guilty (low-quality) arrests. I estimate equation (3) on propensities to make non-guilty and guilty arrests separately. By comparing the effect of peer diversity on the propensity to make high-quality arrests with its effect on the propensity to make low-quality arrests, I can infer an effect of peer diversity on arrest quality separate from the influence of assignments, as crimes committed in some locations and times may be easier to arrest and prosecute than others. In Table 5, Columns (1)-(2) and (5)-(6) display the results for

[36] As previously described, a guilty outcome means the arrest was associated with any guilty finding, and a non-guilty outcome means the arrest was not associated with any Cook County court case, resulted in only dropped charges, or had charges only found not guilty in court. Cases which had no final disposition or closed date in the data set are considered incomplete/open and are neither guilty nor non-guilty. Police officers do influence the initial charges against the arrestee and provide evidence and testimony to prosecutors and defendants, though their time in front of a jury or judge is limited particularly given the frequency of plea deals. Furthermore, while officer observables and behavior influence credibility in the eyes of judges, prosecutors, and juries, which alters their ability to make guilty arrests, an officer's cohort diversity is not observable to them and far removed. Additionally, officers may have reputations for being credible or not for prosecutors, meaning whether charges are filed may be a function of officer 'quality' as well. Lastly, a low-quality arrest does not necessarily mean the arrestee was innocent of any crime.

Black low-level arrests and Columns (3)-(4) and (7)-(8) display the results for Black serious arrests, with Columns (1)-(4) being main sample results and Columns (5)-(8) being full sample results, and guilty arrests in even columns and non-guilty arrests in odd columns.

For the main sample, Black and non-Black minority peers have a small effect on officers making guilty low-level arrests, but they have large negative effects on officers' propensities to make non-guilty low-level arrests. To put the effects in context, the guilty / non-guilty rate for low-level Black arrests for main sample officers is 0.25, and similarly the ratio of the respective standard deviations in shrunken main sample officer fixed effects is 0.21. So, a 10pp increase in CSB decreases non-guilty low-level arrests by -0.0165 ($p < 0.05$) and guilty low-level arrests by -0.001 ($p > 0.1$), which corresponds to a guilty/non-guilty ratio decrease of 0.06. This indicates that CSB has a strong positive effect on arrest quality for low-level arrests because non-guilty arrests decline. Similarly, CSN has a small negative effect on non-guilty low-level arrests (-0.01,$p < 0.05$) and virtually no effect on guilty arrests.

For serious arrests of Blacks, only Black peers have any economically or statistically significant effect, with the coefficient on CSB for guilty serious arrests being 0.005, $p < 0.01$ and a noisy positive effect on non-guilty serious arrests, -0.002, $p > 0.1$. Given that the guilty/non-guilty ratio for serious arrests of Blacks is 0.33, this indicates Black (and non-Black minority, based on noisy coefficients) peers have a very small but positive effect on arrest quality for serious crimes.

This pattern of results is consistent in the full sample as well. Focusing on the interaction of age and minority status, Columns (5)-(8) display these results. Column (5) and (6) indicate that minority and mid- and old-white peers have a significant negative effect on non-guilty low-level arrests of Blacks and effects on guilty low-level arrests of Blacks more than order of magnitude smaller– indicating an increase in arrest quality. As in the previous section, the strongest negative effects are due to the share of older minority peers. For serious arrests, the effects are all statistically and economically insignificant indicating no significant change in arrest quality or quantity.

Low-quality arrests, however, are heterogeneous, as an arrest can result in a non-guilty outcome if 1. it results in a finding of not guilty (usually a judge ruling), 2. the case is dropped by the prosecutor or dismissed by the judge, or 3. it is missing from the court data.[37] In order to determine which of these is driving the low-quality results, I decompose low-quality, low-level arrests of Black civilians and redo my analysis on each sub-type. The results are displayed in Table B.3. The results show that dismissed/dropped cases and not

[37] An arrest may be missing either because the officer later decided not to charge the individual (e.g. 'drunk tank' arrests or protesters), the individual was immediately sent to a diversion program, or due to matching error.

guilty cases are driving the results, with no effect on missing cases, for both the main and full samples. This indicates that reduction is driven by cases with, for example, insufficient evidence or credible testimony by the officer, rather than matching error or the officer choosing to detain but not charge the individual with a sufficiently serious crime. This is consistent with officers being more discerning in their low-level arrests and thereby improving their average arrest quality.

5 Robustness

In this section, I discuss a variety of additional analyses that include alternative samples and specifications to test the robustness of the main results for both the main and full samples. Table 6 presents the main sample robustness tests and Table 7 presents those for the full sample, with each main sample test in Columns (1)-(7) being replicated for the full sample in the analogous column. Columns (8)-(11) in Table 6 present additional tests that are relevant or feasible for the main sample only. Due to the computational intensity of computing fixed effect standard errors, the fixed effects will be unshrunken except for Columns (7) and (11). Finally, I test the robustness of my results to common issues in peer effect studies.

5.1 Discrete Outcomes in First Stage

Arrests in a shift are count data, and the distribution of their frequency, as expected, fits a Poisson distribution. As such, I re-estimate the first stage (equation (2)) using a Poisson regression.[38] This model is potentially more reflective of the true data generating process, and environment and individual officer fixed effects likely contribute to arrests in a non-linear fashion. However, unlike the linear model, the estimates are not directly interpretable and fewer individual fixed effects can be recovered.[39] I use the recovered fixed effects in my second stage, and, as shown in Column (1) of Tables 6 and 7, the results are qualitatively similar to those of the main results, though not directly comparable due to the non-linearity of the model.

While the Poisson first stage is designed to more closely match the distribution in the data, a second concern may be that the results are driven by the skewed nature of the arrest

[38]Specifically,

$$\mathbb{E}[Arrest^k_{it}|\theta^k_i, \gamma^k_{brsw_t}, V_{it}] = exp(\theta^k_i + \gamma^k_{brsw_t} + \beta^k V_{it}) \tag{4}$$

[39]The estimation is performed using the R package 'fixest' and an algorithm used to efficiently estimate fixed effects in maximum likelihood models (Bergé (2018)). As the data are not overly dispersed, a negative binomial regression is not necessary.

data: most shifts have no arrests at all while very few have many. To test this, I re-estimate equation (2) as a linear probability model (LPM) with the dependent variable being if any arrest of type k was made by officer i during their shift. The results are displayed in Column (2) of Tables 6 and 7, and the estimates are very similar to the main results. These tests indicate that the results were not driven by either the reliance on a linear model in the first stage nor the skewed distribution of arrests per shift.

5.2 Alternative Samples and Controls

As multiple officers can be listed on a single arrest, this means some arrests are double counted in my analysis. This may be an issue if cohort share minority influences assignments in which only single officer arrests generally occur. To check the robustness of my results against this issue, I re-estimate the first stage only counting arrests for the first arresting officer. Column (3) in Table 6 displays the results of equation (3) on these recovered fixed effects, and they are similar to the main results but slightly smaller, as expected.

In the main results, the only control included is a polynomial of officer tenure as other factors, such as crime rates or shift duration may be endogenous to the officer's activity during their shift. As a robustness check, I re-estimate officer fixed effects with polynomials of local crime and watch duration during the officer's shift, and I use these officer fixed effects to use as the outcome in equation (3); I find the results are qualitatively similar (see Columns (4)).

While the assignment fixed effects control for significant amounts of heterogeneity in assignment, they do not control for the type of assignment an officer has, e.g. in car, on foot, or on bike patrol (though this is often determined by their beat and shift). To ensure that differential assignment types are not contaminating my results, I re-estimate equation (2) using the 91.9% of panel observations where an officer is in a car. The results are displayed in Columns (5) and are highly similar to the main results.

We used serious and low-level crime types in order to distinguish between types of arrests; however, we want to ensure that the results are not driven by this classification. I re-categorize the arrests based on the FBI index (serious: murder, rape, robbery, aggravated assault, burglary, theft, motor vehicle theft, and arson) and non-index (low-level: all others) classification and exclude warrant arrests as the exact crime type is not known. Fixed effects for Black index and non-index arrests are obtained by re-estimating equation (2). The results in Columns (6) are consistent with the main results.

Finally, with additional information about officers, I am able to add more controls about both officer and cohort compositions in order to test the robustness of the main results. These additional controls include officer level and cohort shares of Spanish speaking, female, high education (bachelors or above), and military-experienced officers as well as cohort mean start age. Columns (7) illustrates that the results are not significantly impacted.

5.3 Main Sample Only Tests

As previously discussed, the main sample cohorts have multiple advantages which also allow for more detailed exploration of effects, including additional robustness checks that are infeasible with full sample cohorts. First, the potential for non-Black officers to be sent to different assignments due to higher shares of Blacks in their cohorts may negatively bias my results (increasing magnitude) as high cohort diversity may lead to low fixed effects solely due to assignments (though the high dimensional working environment controls attempt to solve this issue). As a robustness test, I study the subset of new officers exposed to very high crime areas (above 75th percentile in violent index crime). As shown in Column (8) in Table 6, the 'high crime' new officers display similar results as the whole sample.[40]

Second, there may be a concern that new officers are not only assigned based on race and the racial composition of their cohort, but also based on their preferences and unobservables with respect to policing as well. I repeat my analysis on main sample officers during their probationary periods—which alleviates this issue as they have no actual policing experience upon entering this period– and I include controls for field training evaluator characteristics. Column (9) in Table 6 display these results, with effects qualitatively similar to the main results.[41]

Next, I repeat my analysis using a restricted sample similar to that of Ba et al. (2021), which studied differences in policing outcomes between officers of different races and gender. Their sample differs in two main ways: first, it is restricted to 2012-2015 and excludes watch 4; second, it the assignment fixed effects used are more granular, controlling for the interaction between year-month, day of the week, shift, and exact beat code ('MDSB'), where as I interact assignment role with a truncated beat code.[42] I re-estimate the first stage on the restricted sample and control for assignment more granularly with MDSBs. Column (10) in Table 6 present the results which are similar to those of the main sample.

[40]This analysis is feasible in the main sample but not full sample as there is significant heterogeneity across exams in local crime rates and initial assignments, making it infeasible to decide on a sensible "high crime" cut off across 8 years of cohorts.

[41]Probationary field training officer data is only available fully for the main sample officer cohorts.

[42]The main analysis has other minor differences with their approach. I do not filter out officers for having additional information codes (e.g., indicating injury, training, union business, etc.) during their shifts as this information is not available for my full sample of assignment data. Naturally, this re-estimation is not feasible for Exam 2013 officers, which all start as full officers after this data ends in 2015.

Lastly, while eligible applicants are permitted to enter the academy when their lottery number is drawn (and passing further tests), it is not required. A potential recruit may choose not to join the academy for a variety of reasons if their number is called. While all the estimates in this paper are conditional on uptake (i.e., only relevant for people who end up becoming officers), the composition of cohorts may be influenced by selection into the academy due to different start dates (though Appendix A.2 provides evidence against significant selection).[43] I test this by estimating equation (3) on the subset of cohorts which started within 5 months of the initial main sample cohort, i.e. starting between July and December of 2012. Column (11) in Table 6 displays these results. The results for low-level arrests are qualitatively similar to those in the main results and the effects on serious arrests are insignificant. For further evidence, Figure B.6 displays how the coefficients of interest change as more cohorts are added.[44]

5.4 Measurement Error, Exclusion Groups, and Inference

Finally, there are three common issues with peer effect studies, particularly in settings that utilize similar assignment mechanisms as this paper. First, as discussed in Angrist (2014), building off of Acemoglu and Angrist (2000), and in Feld and Zölitz (2017), peer effects can be over-estimated due to measurement error. In order to determine if measurement error in cohort composition is driving the magnitudes of my results, I follow Carrell, Hoekstra, and Kuka (2018) by adding measurement error to cohort racial compositions. Figure B.7 displays the results of adding increasing amounts of measurement error to the main specifications for the main sample and full sample. For the main sample (Panel A), increasing measurement error attenuates the main results, while in Panel B increasing measurement error modestly attenuates the main results. The Panel B results, which intersect age and minority status, are less attenuated by adding measurement error to race because no error is added to peer age, which also has an effect. Overall, this suggests that measurement error is unlikely to be driving the results.

Second, Angrist (2014) also shows that there is a negative mechanical correlation between an individual and their peer composition, and this mechanical relationship may be driving the results. For the main specifications, I have used leave-out-means for peer com-

[43]Accessing lists of individuals who did not enter the academy is not possible due to data privacy restrictions. However, for policy relevance, the effect of peer composition on the population of individuals who end up becoming police officers is likely more important than the effect of peer composition on the population on the margin of becoming police officers.

[44]This is not feasible for the full sample because while about 40% of recruits in the main sample start within the first 5 months, it takes significantly longer for a sizable portion of Exam 2013 recruits to enter the academy, and there are only 5 cohorts in the Exam 2006 data set spread out across 2 years.

positions, so to determine if the mechanical correlation is driving the results I distinguish between the individuals being influenced from the peers who are doing the influencing (Angrist and Lang (2004), Imberman, Kugler, and Sacerdote (2012), Carrell, Hoekstra, and Kuka (2018)). Table B.4 presents the results when excluding the peer group in question from the regression. Columns (1) and (2) present the effect of minority peers on white officers (excluding minority officers), Column (3) presents the effect of older peers on young and mid-aged officers (excluding older officers), and Column (5) presents the effect of female peers on male officers (excluding female officers); Columns (4) and (6) add interactions with officer race being white (see Section 6). The results are generally consistent with the results in Sections 4 and 6 (though generally noisier due to smaller samples), indicating that the results are robust to excluding the effecting officers.

Lastly, traditional inference techniques do not necessarily apply to many (quasi-)experimental designs, particularly peer effects studies where inter-group variation is a result of finite-sample bias. Recent peer effect studies use randomization inference to construct p-values for estimates (Carrell, Sacerdote, and West (2013), Caeyers and Fafchamps (2016), Carrell, Hoekstra, and West (2019)), consistent with the guidance in Athey and Imbens (2016). I construct p-values using a randomization inference method which provides a distribution of estimates under the null hypothesis that there is no effect of peer composition on outcomes (see Appendix A.7 for more details). I construct 1,000 placebo cohorts with recruits randomly re-assigned at the exam level, then I re-compute cohort compositions and re-run the main regressions, and two-sided p-values are computed by ranking the coefficient in the main results within the distribution of placebo coefficients (in absolute value). As shown in Figure 2, the randomization inference p-values are generally smaller than those in the main results.

6 Mechanisms

There are multiple potential mechanisms underlying these results: positive interracial socialization, peer preferences, and negative interracial interactions. Overall, the evidence presented in this section is most consistent with the positive interracial socialization and the influence of peer preferences. I begin by elaborating on these two mechanisms, then I present evidence for them and discuss their implications. Finally, I discuss alternative explanations, such as peer race being correlated with other influential characteristics or instructors being influenced by the recruits they teach. To do so, I leverage the fact that additional peer characteristics, such as education and gender, are effectively randomly assigned by the lottery number system.

6.1 Primary Mechanisms

First, positive interracial socialization operates through contact between whites and minorities. During the academy, whites and minorities may become friends, causing a reduction in racial bias or prejudice among whites (Boisjoly et al. (2006), Carrell, Hoekstra, and West (2019)). Evidence consistent with positive interracial socialization includes the effect of minority peers on whites being larger than the effect of minority peers on minorities.

Second, officers may adopt the preferences of those around them due to shifts in culture, social identity, or personality (Akerlof and Kranton (2000), Anwar, Bayer, and Hjalmarsson (2019), Golsteyn, Non, and Zölitz (2021)), meaning each recruit's preferences are influenced by the composition of their peers' preferences. For example, having more peers who prefer aggressive policing will cause an officer to police more aggressively in the future. While I can only observe officer behavior, I use arrest propensities as a proxy for officer preferences. Evidence consistent with the preference effect includes all officers being influenced by minority peers and the peer effect of a group being larger if that group, on average, polices less aggressively (e.g., the effect of female peers should be greater than male peers, as women tend to police less aggressively).

To determine the mechanisms behind the results, I first re-estimate equation (3) with additional terms interacting an officer being white with the variables of interest (minority cohort shares). Table 8 displays the results in Column (1) for the main sample and (2)-(6) for the full sample. I focus solely on low-level arrests of Black civilians because the main results for serious arrests were less precise and not economically significant. The net effect on whites can be calculated as the base coefficient (e.g., Cohort Share Black) added to the interaction term (e.g., White x Cohort Share Black), while the net effect on minorities is just the base coefficient (e.g., Cohort Share Black). Figure B.8 displays net effects on low-level and serious arrests visually for easier interpretation. For the main sample (Column (1)), the effects of Black and non-Black minority peers on white officers' propensities to arrest Blacks for low-level crimes are economically larger. There are multiple reasons why non-Black minority peers may reduce white officers' aggressive policing of Blacks: having more non-Black minorities (fewer whites) may facilitate Black-white contact and friendship; or contact with Hispanics, for example, may reduce bias against Blacks through the secondary transfer effect, documented in social psychology (Pettigrew (2009)). Overall, the point estimates are consistent with both interracial socialization (whites being affected the most) and the influence of peer preferences (minorities are influenced as well); however, they are relatively imprecise, likely due to the small sample size.

For more precise estimates, I use the full sample for the remaining analyses. In Column

(2), I include interactions for officers being white on their cohort shares of young, mid, and old minorities and mid/old whites. The effect of older minority peers on minorities is the largest (-0.166, $p < 0.01$), and the effect size decreases as the age group decreases. As older officers tend to police less aggressively (see Table B.1), this is consistent with the effect of peer preferences. By contrast, white officers' propensity to arrest Blacks for low-level crimes are larger and negative for all minority age groups, consistent with positive interracial socialization.

I further explore these mechanisms in remaining columns in Table 8. First, it is important to determine more precisely the relationship between age and minority race. In the main sample, Black peers had the strongest effect, and given that Black recruits tend to be older than non-Black recruits, we want to disentangle the race and age effects to see if the age group results are actually due to Black peers. In Column (3), I regroup minorities into Black and non-Black minorities and interact them with age groups young/mid and old (i.e., cohort shares for each combination of $\{\textit{Black}, \textit{Non-Black Minority}\} \times \{\leq 32, > 32\}$). The results in Column (3) indicate that although Black peers of all ages decrease arrest propensities, the largest and most precisely estimated effects are of older Black peers. The amplifying effect of age is also true for non-Black minority peers as well. So, while older minorities drive the effect of minorities on arrest propensities, older Black peers have an effect almost twice the size of older non-Black minority peers.

In Column (4), I add interactions with white officers, and the net effects are displayed in Figure B.8 for easier interpretation. As in Column (3), older Black and non-Black minority peers have the largest effects on minority officers relative to younger minority peers, with Black peers having a larger effect than non-Black minority peers. The effect of Black and non-Black minority peers on whites is larger than their effect on minorities, particularly for the effect of younger Black and non-Black minority peers. These results are further evidence for interracial socialization influencing whites, and peer preferences causing shifts in all officers' future behavior on the job. Notably, in the full sample, even mid-age and older white peers have a significant effect on all officers, but no additional effect on whites, consistent with the fact that older white officers prefer less aggressive policing relative to younger whites (see Table B.1) but have no additional bias reduction effect on whites.

Gender is also correlated with minority status and age, so it may be that female minority peers are actually driving the effects. In Column (5), I add an additional control for cohort shares of white female and minority female recruits, and Column (6) adds the white officer interaction term. While the effects of minority and white peers by age group are similar to those of the Table 4, the additional coefficients indicate that female peers tend reduce arrest propensities, with white female peers having a small and not statistically significant effect (-

0.049, $p > 0.1$) and minority female peers having an economically and statistically significant effect (-0.14, $p < 0.05$). This effect is in line with the fact that female officers tend to police Blacks less aggressively relative to male officers within race (see Table B.1 and Ba et al. (2021)). Adding the race-white interaction term in Column (6) has little effect on the main coefficients, but it reduces the effect of female peers on minority officers. The coefficients suggest that female minority peers have a statistically and economically significant effect on all officers with a very small additional effect on white officers. However, white female peers have a small and very imprecise effect on minority officers and a relatively large effect on white officers. This pattern is consistent with inter-gender socialization being more common within race, and thus white females (who are less aggressive than white males) influencing their white male peers. While the interactions terms are imprecise, they are consistent with peers with propensities for less aggressive policing influencing all officers to police less aggressively.

There are three main implications. First, the evidence for interracial socialization reducing aggressive policing by whites highlights peer effects as one way in which departmental diversity alters officer behavior. Second, while policies such as procedural justice training have not produced significant changes in officer behavior (Roth and Sant'Anna (2021)) and common-place diversity trainings can be ineffective or counterproductive (Chang et al. (2019), Dobbin and Kalev (2016)), these results indicate that having diverse peers is an effective way of reducing aggressive policing. Finally, the effects of peer preferences are quite large. This indicates that while demographic representation is important, hiring diverse officers who have preferences for more aggressive policing (e.g., young and male) may actually be detrimental, as their peers may police more aggressively as a result.

6.2 Alternative Explanations

Although the evidence has thus far been consistent with positive interracial socialization and peers' preferences influencing officer behavior, there are alternative explanations. First, racial or age group classification could be correlated with other characteristics such as education or military experience. Education, for example, has been shown to be an important moderating factor in peer effect of diversity, as Carrell, Hoekstra, and West (2019) finds that students assigned to peer groups with more mid- and high-performing Black students had increased future interactions with Blacks. Second, instructors may be influenced by class composition (Lavy and Schlosser (2011)), thus changing how officers learn to police. For example, instructors with more Black students may treat them more respectfully, leading to officers policing Blacks less aggressively on the job. If education or military experience or instructors, and not peer racial composition, are driving the results, then the policy implications would be significantly different. Naturally, the evidence thus far completely rules out negative interracial interactions leading to animus as a mechanism.

For education and military experience, I explore these potential explanations in Table B.5 focusing on low-level arrests of Blacks using the full sample. Column (1) of Table B.5 displays the results from including additional controls for cohort share white and minority with high-education (a Bachelors or above) before the academy, and Column (2) adds the white officer interaction to these controls. In both cases, the variables of interest are unaffected. Peer education of whites or minorities does not seem to have any effect nor does it strongly influence the variables of interest of Table B.5. This is expected if officers are influenced by peer preferences because officer education is not associated with lower propensities for aggressive policing (see Table B.1). The interaction term in Column (2) for cohort share high education minority and officer race white is small, positive, and not statistically significant, while the interaction with cohort share high education white is positive and statistically significant. This indicates education is not mediating interracial socialization (i.e., whites are not only influenced by high education minority peers).

The same exercise is repeated for military experience in Columns (3) and (4). White military peers have a negative effect (-0.19, $p < 0.01$) and minority military peers have a positive effect (0.161, $p = 0.069$) on arrest propensities, which causes the effect of non-military minority peers of all age groups to be more precisely estimated and significantly larger. However, given that 97% of full sample recruits have military experience, the size of the effect is largely overstated. By comparison, the interaction terms in Column (4) are very small. Overall, being in the military attenuates peer effects, as military minority peers have smaller negative effects and military white peers have larger negative effects, but these effects are small in reality due to high military shares. From this and the education results, we can conclude that neither peer education nor military experience are driving the results, and they do not confound the previous evidence for the interracial socialization and peer preference mechanisms.

Finally, we turn to the possibility that peer effects of diversity are operating by influencing instructors, i.e., instructors behave differently in the presence of more minorities thus changing how recruits learn to police. While this cannot be directly tested as instructor identities are not known, there are multiple reasons why it is unlikely. First, this would likely be a very subtle effect, particularly given instructors are generally veteran police officers or related experts in law enforcement, who may not be particularly sensitive to class composition. This effect would likely also operate in courses related to conduct or ethics. Yet this training is quite minimal in the CPD curriculum, consistent with surveys on police training across the U.S. (Cohen (2021)), implying instructor effects would be even more subtle.[45]

Furthermore, we would expect that if class composition is influencing instructors it would be strongest in small classes. For example, an instructor may be more likely to notice a 10pp increase in share Black in a class of 20 recruits relative to a class of 70. In Appendix A.8, I test the effect of small class shifts, and the results indicate that it is unlikely that the effects of small-group composition are driving the results, which does not support instructor effects. Notably, the inclusion of Field Training Officer characteristics in Column (9) of Table 6 did not reduce the effect of minority peers during the probationary period– which would have been evidence of instructor effects during the probationary period. Overall, instructor effects, if present, are likely subtle and not the primary mechanisms driving the results.

7 Conclusion

In this paper, I document the effect of minority peers in the police academy on police officers' future arrest quantity and quality. Higher shares of minority peers significantly reduce officers' future propensity to arrest minorities for low-level crimes. This effect is driven by a decline in low-quality (not resulting in a guilty finding) arrests of Blacks ($> 80\%$ of arrests), implying an increase in average arrest quality. Importantly, I find that minority peers have a small positive or null effect on arrests for serious crimes, implying officers are not reducing effort toward combating threats to public safety or property.

I find evidence consistent with two main mechanisms. First, white officers are most strongly influenced by minority peers of all age groups, consistent with interracial socialization reducing racial biases. Second, minority peers also reduce minority officers' propensity to police aggressively, though this is strongest for older and female minority peers, consistent with a shift in enforcement behavior due to academy peers' preferences. Notably, the bias reduction effect is smaller than the peer preference effect. I do not find evidence for instructor effects or negative interracial contact. Furthermore, the effects are present for at least 4 years after the academy ends.

These results indicate that beyond minority status, race, gender, and age are all important factors which influence peer effects. In general, officers who police less aggressively themselves (e.g., Black, female, and older officers) cause larger shifts in their peers' propensities to police aggressively. If arrest propensities are indicative of officer preferences, then these results indicate that policy changes that result in more recruitment of minorities who prefer more aggressive policing may have self-defeating effects. For example, lowering mini-

[45] Cohen (2021) finds that on average only about 1% of training hours were dedicated to ethics, cultural competency, communication, or procedural justice each.

mum start ages in order to have more minority applicants may lead to increased recruitment of younger male minorities, which can nullify the effect of increased racial diversity due to the effect of increased peer preferences for aggressive policing. These results suggest that additional characteristics should be considered to improve departmental hiring policy, though this is ground for future research.

Overall, these results are generally consistent with the existing literature in economics and psychology on peer diversity in environments significantly different from a police academy. As the CPD and other major police departments all have sizable minority shares, the experiences of CPD officers are likely similar to those in other large cities. Furthermore, while Chicago is a single city, it provides an ideal environment because it is a sufficiently large department in a diverse city, which has a long-standing lottery system for the academy and maintains high quality data. Thus, the policy implications of these findings are far reaching and promising for improving policing. The inclusion of minority officers can result in persistent effects on their peers, reducing over-policing of low-level offenses while not reducing propensities to make arrests for more serious crimes, thereby potentially improving police-community relationships. Importantly, officers' arrest quality increases with peer diversity, so increasing departmental diversity through the recruitment of more minority officers can result in fewer wasted public resources, fewer individuals put under undue burdens, and fewer separated families.

8 References

Acemoglu, Daron, and Joshua Angrist. 2000. "How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual* 15 (January): 9–59.

Aendos. 2015. "Chicago Police 2016." *Police Forums & Law Enforcement Forums @ Officer.com.* https://forum.officer.com/forum/local-discussion-groups/u-s-states/illinois/206320-chicago-police-2016/page138.

Agan, Amanda Y., Jennifer L. Doleac, and Anna Harvey. 2021. "Misdemeanor Prosecution." w28600. National Bureau of Economic Research.

Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth. 2021. "Killer Incentives: Rivalry, Performance and Risk-Taking Among German Fighter Pilots, 1939–45." *The Review of Economic Studies*, December.

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics* 130 (2): 759–803.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–53.

———. 2010. *Identity Economics : How Our Identities Shape Our Work, Wages, and Well-Being.* Princeton: Princeton University Press.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics*, Special Section articles on "What determined the dynamics of labour economics research in the past 25 years? Edited by Joop Hartog and and European Association of Labour Economists 25th Annual Conference, Turin, Italy, 19-21 September 2013 Edited by Michele Pellizzari, 30 (October): 98–108.

Angrist, Joshua D., and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review* 94 (5): 1613–34.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics* 127 (2): 1017–55.

———. 2019. "Politics in the Courtroom: Political Ideology and Jury Decision Making." *Journal of the European Economic Association* 17 (3): 834–75.

Ater, Itai, Yehonatan Givati, and Oren Rigbi. 2014. "Organizational Structure, Police Activity and Crime." *Journal of Public Economics* 115 (July): 62–71.

Athey, Susan, Dean Eckles, and Guido W. Imbens. 2018. "Exact p-Values for Network Interference." *Journal of the American Statistical Association* 113 (521): 230–40.

Athey, Susan, and Guido Imbens. 2016. "The Econometrics of Randomized Experiments."

arXiv:1607.00698 [Econ, Stat], July.

Austen-Smith, David, and Roland G. Fryer Jr. 2005. "An Economic Analysis of 'Acting White'." *The Quarterly Journal of Economics* 120 (2): 551–83.

Ba, Bocar A., Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. "The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science* 371 (6530): 696–702.

Baker, Sara, Adalbert Mayer, and Steven L. Puller. 2011. "Do More Diverse Environments Increase the Diversity of Subsequent Interaction? Evidence from Random Dorm Assignment." *Economics Letters* 110 (2): 110–12.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. 2010. "Social Identity and Preferences." *The American Economic Review* 100 (4): 1913–28.

Bergé, Laurent R. 2018. "Efficient Estimation of Maximum Likelihood Models with Multiple FIxed-Effects: The r Package FENmlm." *CREA Discussion Papers*, no. 13: 39.

Billings, Stephen B., Eric Chyn, and Kareem Haggag. 2021. "The Long-Run Effects of School Racial Diversity on Political Identity." *American Economic Review: Insights* 3 (3): 267–84.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2013. "Under Pressure? The Effect of Peers on Outcomes of Young Adults." *Journal of Labor Economics* 31 (1): 119–53.

Blanes i Vidal, Jordi, and Tom Kirchmaier. 2018. "The Effect of Police Response Time on Crime Clearance Rates." *The Review of Economic Studies* 85 (2): 855–91.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *The American Economic Review* 96 (5): 1890–1905.

Brenøe, Anne Ardila, and Ulf Zölitz. 2020. "Exposure to More Female Peers Widens the Gender Gap in STEM Participation." *Journal of Labor Economics* 38 (4): 1009–54.

Burns, Justine, Lucia Corno, and Eliana La Ferrara. 2015. "Interaction, Prejudice and Performance. Evidence from South Africa." *Working Paper*, February, 52.

Caeyers, Bet, and Marcel Fafchamps. 2016. "Exclusion Bias in the Estimation of Peer Effects." w22565. Cambridge, MA: National Bureau of Economic Research.

Card, David, and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100 (1): 1–40.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review* 108 (11): 3377–3415.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2019. "The Impact of College Diversity on Behavior Toward Minorities." *American Economic Journal: Economic Policy* 11 (4): 159–82.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3): 855–82.

Chang, Edward H., Katherine L. Milkman, Dena M. Gromet, Robert W. Rebele, Cade Massey, Angela L. Duckworth, and Adam M. Grant. 2019. "The Mixed Effects of Online Diversity Training." *Proceedings of the National Academy of Sciences* 116 (16): 7778–83.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics* 126 (4): 1593–1660.

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.

Chicago_mwk. 2010. "Chicago Police Academy 2010." *Police Forums & Law Enforcement Forums @ Officer.com*. https://forum.officer.com/forum/local-discussion-groups/u-s-states/illinois/141254-chicago-police-academy-2010.

Close, Billy R., and Patrick L. Mason. 2007. "Searching for Efficient Enforcement: Officer Characteristics and Racially Biased Policing." *Review of Law and Economics* 3 (2): 263–322.

Cohen, Galia. 2021. "Public Administration Training in Basic Police Academies: A 50-State Comparative Analysis." *The American Review of Public Administration* 51 (5): 345–59.

Coviello, Decio, and Nicola Persico. 2015. "An Economic Analysis of Black-White Disparities in the New York Police Department's Stop-and-Frisk Program." *The Journal of Legal Studies* 44 (2): 315–60.

Cox, Robynn, Jamein P Cunningham, and Alberto Ortega. 2021. "The Impact of Affirmative Action Litigation on Police Killings of Civilians." Working Paper.

CPD. 2011. "Personnel Transfer and Assignment Procedures – (FOP)." http://directives.chicagopolice.org/directives/data/a7a57be2-12bcf25e-31612-bcf2-5ebc1c9f5d96947f.html?hl=true.

———. 2017. "CPD 2017 FAQ."

———. 2018. "Field Training and Evaluation Program." http://directives.chicagopolice.org/directives/data/a7a57be2-1294231a-bf312-942c-e1f46fde5fd8c4e8.html?hl=true.

———. 2020. "Education and Training Division (ETD) |Chicago Police Department." https://home.chicagopolice.org/about/specialized-units/education-and-training-division-etd/.

Csardi, Gabor, and Tamas Nepusz. 2005. "The Igraph Software Package for Complex Network Research." *InterJournal* Complex Systems (November): 1695.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.

Dobbin, Frank, and Alexandra Kalev. 2016. "Why Diversity Programs Fail." *Harvard Business Review* 94 (7).

DOJ. 2016. "Advancing Diversity in Law Enforcement Report (October 2016)." Report. The United States Department of Justice: DOJ Civil Rights Division; Equal Employment Opportunity Commission.

Donohue III, John J., and Steven D. Levitt. 2001. "The Impact of Race on Policing and Arrests." *The Journal of Law & Economics* 44 (2): 367–94.

Feld, Jan, and Ulf Zölitz. 2017. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics*, January.

feredeathpsn. 2017. "Chicago Police Lottery Number." Reddit Post. *R/Police*. www.reddit.com/r/police/comments/6k5fun/chicago_police_lottery_number/.

Ferguson, Joseph M, and Deborah Witzburg. 2021. "EVALUATION OF THE DEMOGRAPHIC IMPACTS OF THE CHICAGO POLICE DEPARTMENT's HIRING PROCESS." Report. City of Chicago: Office of the Inspector General.

Fisher, R. A. 1925. "Theory of Statistical Estimation." *Mathematical Proceedings of the Cambridge Philosophical Society* 22 (5): 700–725.

Fryer, Roland G., and Paul Torelli. 2010. "An Empirical Analysis of 'Acting White'." *Journal of Public Economics* 94 (5): 380–96.

Garner, Maryah, Anna Harvey, and Hunter Johnson. 2019. "Estimating Effects of Affirmative Action in Policing: A Replication and Extension." *International Review of Law and Economics*, November, 105881.

Gaure, Simen. 2013a. "Lfe: Linear Group Fixed Effects." *The R Journal* 5 (2): 104.

———. 2013b. "OLS with Multiple High Dimensional Category Variables." *Computational Statistics & Data Analysis* 66 (October): 8–18.

Golsteyn, Bart H. H., Arjan Non, and Ulf Zölitz. 2021. "The Impact of Peer Personality on Academic Achievement." *Journal of Political Economy* 129 (4): 1052–99.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–41.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman. 2009. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence." *The*

Economic Journal 119 (540): 1243–69.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The Heavy Costs of High Bail: Evidence from Judge Randomization." *Journal of Legal Studies* 45 (2): 471–505.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics* 1 (4): 34–68.

Hahn, Jinyong, and Zhipeng Liao. 2021. "Bootstrap Standard Error Estimates and Inference." *Econometrica* 89 (4): 1963–77.

Harvey, Anna, and Taylor Mattia. 2019. "Reducing Racial Disparities in Crime Victimization." *Working Paper*, December, 45.

Hoekstra, Mark, and CarlyWill Sloan. 2020. "Does Race Matter for Police Use of Force? Evidence from 911 Calls." Working Paper 26774. National Bureau of Economic Research.

Holden, Richard, Michael Keane, and Matthew Lilley. 2021. "Peer Effects on the United States Supreme Court." Working Paper.

Holz, Justin E, Roman G Rivera, and Bocar Ba. 2019. "Network Effects in Police Use of Force." Working Paper.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." Working Paper 7867. National Bureau of Economic Research.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102 (5): 2048–82.

Kass, John, and Robert Blau. 1991. "POLICE HIRING LOTTERY LATEST DALEY HEADACHE." *Chicago Tribune*, August, 3.

Keller, Meg. 2015. "Diversity on the Force: Where Police Don't Mirror Communities." Governing.

Kirchmaier, Tom, Stephen Machin, Matteo Sandi, and Robert Witt. 2021. "Joining Forces? Crewing Size and the Productivity of Policing." Working Paper.

Laar, Colette Van, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. "The Effect of University Roommate Contact on Ethnic Attitudes and Behavior." *Journal of Experimental Social Psychology* 41 (4): 329–45.

Lavy, Victor, and Analía Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3 (2): 1–33.

Leifeld, Philip. 2013. "Texreg: Conversion of Statistical Model Output in r to LATEX and HTML Tables." *Journal of Statistical Software* 55 (November): 1–24.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60 (3): 531–42.

Mas, Alexandre. 2006. "Pay, Reference Points, and Police Performance." *The Quarterly Journal of Economics* 121 (3): 783–821.

McCrary, Justin. 2007. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police." *The American Economic Review* 97 (1): 318–53.

Merlino, Luca Paolo, Max Friedrich Steinhardt, and Liam Wren-Lewis. 2019. "More Than Just Friends? School Peers and Adult Interracial Relationships." *Journal of Labor Economics* 37 (3): 663–713.

Michelman, Valerie, Joseph Price, and Seth D Zimmerman. 2021. "Old Boys' Clubs and Upward Mobility Among the Educational Elite." *The Quarterly Journal of Economics*, December.

Miller, Amalia R, and Carmit Segal. 2018. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *The Review of Economic Studies* 86 (5): 2220–47.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78 (381): 47–55.

neverlose357. 2010. "2011 Chicago Police Academy." *Police Forums & Law Enforcement Forums @ Officer.com.* https://forum.officer.com/forum/local-discussion-groups/u-s-states/illinois/161159-2011-chicago-police-academy.

Newman, M. E. J., and M. Girvan. 2004. "Finding and Evaluating Community Structure in Networks." *Physical Review E* 69 (2).

Owens, Emily, David Weisburd, Karen L. Amendola, and Geoffrey P. Alpert. 2018. "Can You Build a Better Cop?" *Criminology & Public Policy* 17 (1): 41–87.

Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology* 49 (1): 21.

———. 2009. "Secondary Transfer Effect of Contact: Do Intergroup Contact Effects Spread to Noncontacted Outgroups?" *Social Psychology* 40 (2): 55.

Pettigrew, Thomas F, and Linda R. Tropp. 2006. "A Meta-Analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology* 90 (5): 751–83.

Pritchard, Paige. 2013. "Do You Have What It Takes to Join the Chicago Police Department?" *Chicago Magazine*, August.

Roth, Jonathan, and Pedro H. C. Sant'Anna. 2021. "Efficient Estimation for Staggered Rollout Designs." Working Paper.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *The Quarterly Journal of Economics* 116 (2): 681–704.

———. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, 3:249–77. Elsevier.

Schindler, David, and Mark Westcott. 2021. "Shocking Racial Attitudes: Black g.i.s in Europe." *The Review of Economic Studies* 88 (1): 489–520.

Shi, Lan. 2009. "The Limit of Oversight in Policing: Evidence from the 2001 Cincinnati Riot." *Journal of Public Economics* 93 (1-2): 99–113.

Sommers, Samuel R. 2006. "On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations." *Journal of Personality and Social Psychology* 90 (4): 597–612.

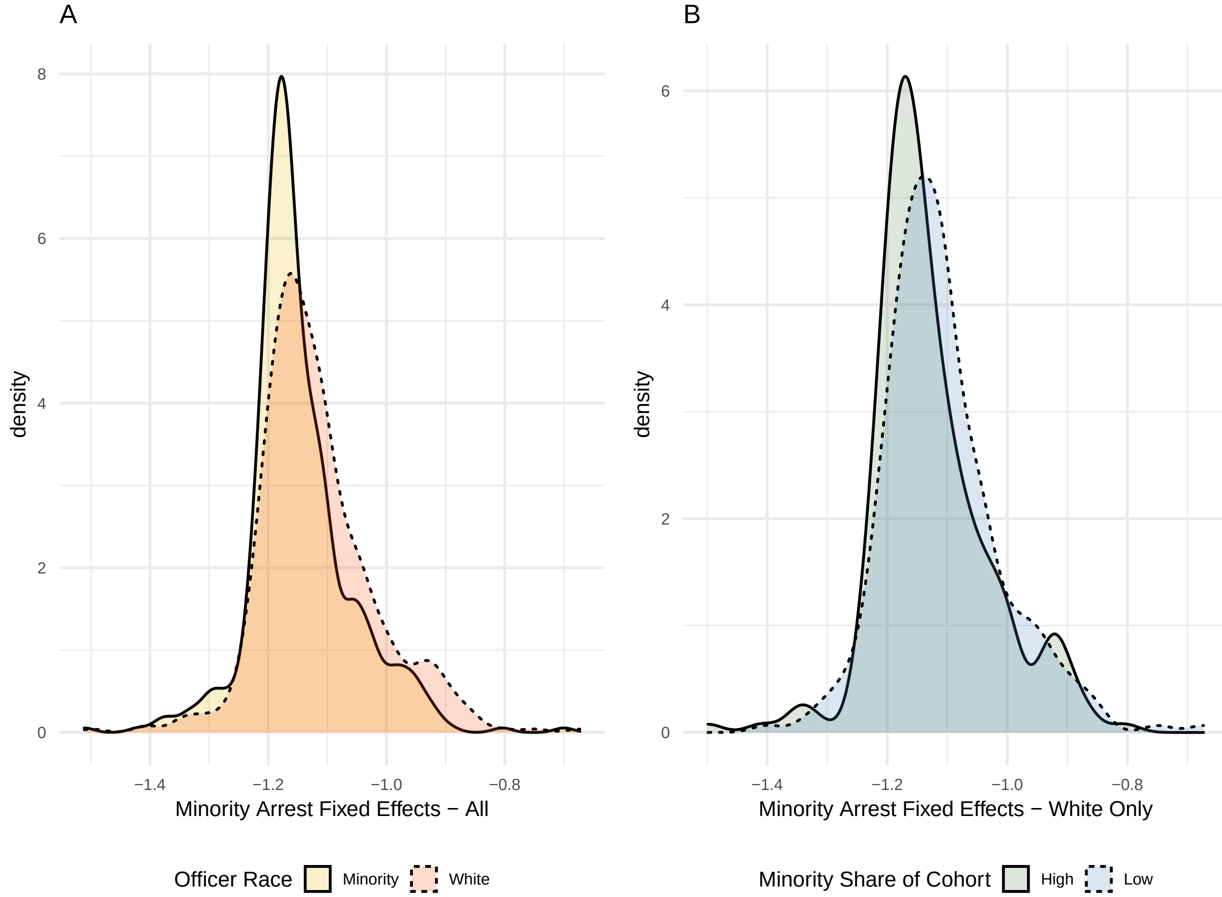
Stevenson, Megan T. 2018. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." *The Journal of Law, Economics, and Organization* 34 (4): 511–42.

Tausch, Nicole, Miles Hewstone, Jared B. Kenworthy, Charis Psaltis, Katharina Schmid, Jason R. Popan, Ed Cairns, and Joanne Hughes. 2010. "Secondary Transfer Effects of Intergroup Contact: Alternative Accounts and Underlying Processes." *Journal of Personality and Social Psychology* 99 (2): 282.

Weisburst, Emily K. 2020. "'Whose Help Is on the Way?' The Importance of Individual Police Officers in Law Enforcement Outcomes." Working Paper.

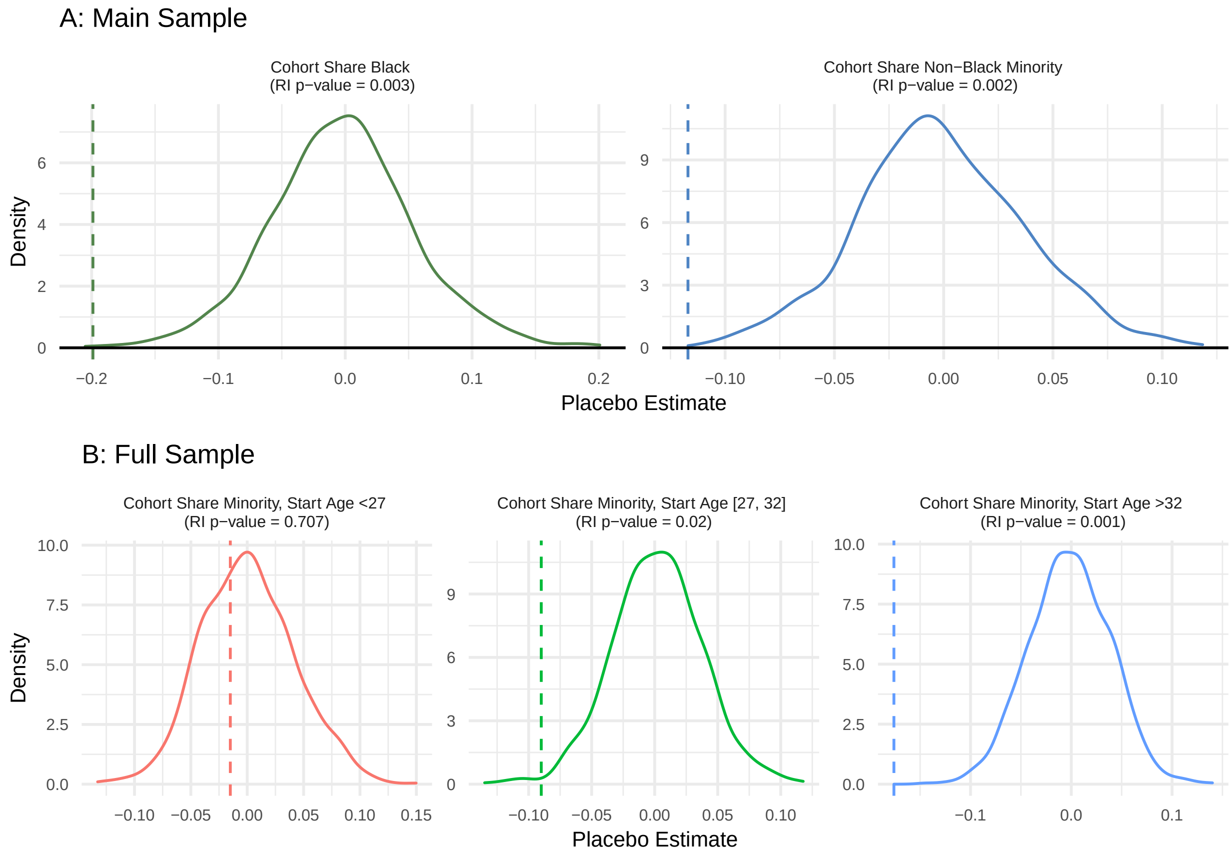
West, Jeremy. 2018. "Racial Bias in Police Investigations." Working Paper.

Figure 1: Distribution of Main Sample Officer Fixed Effects



Note: Figure displays the distributions of main sample officer (Exam 2010) fixed effects recovered from estimating equation (2) with arrests of minorities as the dependent variable. Panel A displays the distributions for white and minority officers separately, and it shows that white officers tend to have higher fixed effects for minority arrests. Panel B displays the distributions of white officers split by on whether they were in a high (at least 50%) or low (below 50%) minority cohort, and it shows that whites in high−minority cohorts tend to have lower fixed effects compared to whites in low−minority cohorts. Displayed fixed effects are generally negative due to the leave−out officer in the first stage having an arrest propensity higher than most main sample officers.

Figure 2: Randomization Distribution of Coefficients (Reassigned Cohorts)



Note: Figure visualizes the distribution of coefficients estimated using 1,000 placebo cohorts to conduct randomization inference, as discussed in Appendix A.7. Placebo cohorts are constructed by re−assigning recruits to cohorts within their exam period, ensuring the same number and size of cohorts. Coefficients are the effects of cohort shares Black and non−Black minorities in the main sample (panel A) and the effects of cohort share minority by age group for full sample officers (panel B) on shrunken officer fixed effects for arresting Blacks for low−level crimes. The dashed vertical lines correspond to the coefficient estimated in the main specification (actual cohorts), and the RI p−value denotes the p−value resulting from a two−tailed test which ranks the magnitude of the actual coefficient among the magnitudes of the 1,000 placebo coefficients. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Coefficients estimated using equation (3) using full controls (specification in Column (13) in Table 3 for main sample and specification in Column (5) in Table (4) for the full sample).

Table 1: Summary Statistics by Sample

	Pooled	Cohorts	Pooled	Cohorts	Pooled	Cohorts	Pooled
	Main Sample (Exam 2010)		Exam 2006		Exam 2013		All Officers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2010	2010	2006	2006	2013	2013	
Male	0.807	0.8	0.808	0.769	0.767	0.778	0.756
Female	0.193	0.2	0.192	0.231	0.233	0.222	0.244
White	0.513	0.506	0.477	0.467	0.489	0.488	0.471
Minority	0.487	0.494	0.523	0.533	0.511	0.512	0.529
Black	0.132	0.13	0.224	0.229	0.126	0.121	0.227
Hispanic	0.312	0.325	0.263	0.273	0.341	0.35	0.261
Asian/Native American	0.0436	0.0396	0.0356	0.0319	0.0448	0.0407	0.041
Birth Year	1982.66	1982.7	1980.28	1980.34	1987.22	1987.15	1975.23
Start Age	30.07	30.16	29.6	29.54	28.18	28.04	28.99
Cohort Size	61.11	54.05	81.75	72	83.63	73.71	-
N	940	21	281	5	1115	17	11391

Note: Table compares the average characteristics of main sample (Exam 2010) officers (Columns (1)-(2)), Exam 2006 (Columns (3)-(4)), and Exam 2013 (Columns (5)-(6)) to all of the officers in the panel data (Column (7)). Column (1) contains the pooled average characteristics over all main sample recruits. Column (2) contains the average characteristics of the cohorts of the recruits in (1), including those recruits that do not appear in the main analysis due to attrition. Columns (3)-(6) replicate (1)-(2) for their respective samples. Column (7) contains the average characteristics of all officers in the daily assignment panel data.

Table 2: Summary Statistics of Main Sample by Cohort Composition

	All	High Minority Cohort	Low Minority Cohort	P-value
	(1)	(2)	(3)	(4)
Cohort Composition				
Share Minority	0.494 (0.064)	0.545 (0.031)	0.438 (0.036)	<0.01
Share Female	0.2 (0.039)	0.192 (0.039)	0.208 (0.04)	0.373
Mean Age	30.16 (0.542)	30.27 (0.565)	30.04 (0.515)	0.333
N. Cohorts	21	11	10	-
Officer Outcomes				
Total Arrests	0.175 (0.45)	0.172 (0.45)	0.178 (0.46)	<0.01
White Arrestees	0.008 (0.09)	0.008 (0.1)	0.008 (0.09)	0.01
Black Arrestees	0.143 (0.41)	0.138 (0.4)	0.148 (0.42)	<0.01
Hispanic Arrestees	0.023 (0.17)	0.025 (0.17)	0.021 (0.16)	<0.01
Total Serious Arrests	0.052 (0.25)	0.051 (0.24)	0.053 (0.25)	0.002
Total Low-Level Arrests	0.123 (0.38)	0.121 (0.38)	0.125 (0.39)	<0.01
Guilty Arrests	0.052 (0.31)	0.05 (0.3)	0.055 (0.32)	<0.01
Index Violent Crime Rate in District	15.4 (7.28)	15.08 (7.44)	15.72 (7.11)	<0.01
Obs	531597	262807	268790	-
Unique Officers	940	484	456	-

Note: Table presents the average cohort compositions at cohort level for main sample cohorts (top panel) and average number of arrests per shift, violent crime rate in average working district, and total observations for main sample recruits as full officers from 2013 to 2018 at the officer level (bottom panel). Columns (2) and (3) divide those recruits by whether or not they were in a cohort with a high (at least 50%) or low (less than 50%) share of minorities. Violent crime rate is determined by the district's average violent index crime rate per 10,000 population (2010 Census) in the month of an officer's assignment. P-values in Column (4) based on two-sided Student's t-Test comparing Columns (2) and (3). Standard deviations are reported in parentheses.

Table 3: Effect of Cohort Diversity on Arrest Propensity - Main Sample

	Avg. Arrests (First 200 Shifts)	Raw Arrest Propensity			Raw Arrest Propensity								Shrunken Arrest Propensity	
	Minority	Minority			Black						Non-Black Minority		Black	
	All	All	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Cohort Share Minority	-0.188^{*}	-0.138	-0.186^{**}	0.048^{*}	-0.196^{**}	0.044^{*}								
	(0.099)	(0.086)	(0.091)	(0.025)	(0.097)	(0.024)								
Cohort Share Black							-0.314^{**}	0.097^{***}	-0.256^{**}	0.080^{***}	-0.058	0.017^{***}	-0.199^{**}	0.028^{**}
							(0.148)	(0.028)	(0.117)	(0.024)	(0.036)	(0.006)	(0.098)	(0.013)
Cohort Share Non-Black Minority							-0.193^{**}	0.043^{**}	-0.153^{**}	0.033^{*}	-0.040	0.009^{***}	-0.117^{*}	0.007
							(0.098)	(0.021)	(0.074)	(0.018)	(0.026)	(0.003)	(0.061)	(0.009)
Black					-0.030^{***}	-0.006^{**}	-0.030^{***}	-0.006^{*}	-0.023^{***}	-0.005^{*}	-0.007^{***}	-0.001	-0.021^{***}	-0.004^{**}
					(0.008)	(0.003)	(0.008)	(0.003)	(0.007)	(0.003)	(0.001)	(0.001)	(0.006)	(0.002)
Hispanic					-0.016^{***}	-0.003^{*}	-0.017^{***}	-0.003^{*}	-0.016^{***}	-0.003^{**}	-0.001	0.000	-0.014^{***}	-0.003^{**}
					(0.006)	(0.002)	(0.006)	(0.002)	(0.006)	(0.002)	(0.001)	(0.001)	(0.005)	(0.001)
Asian/Native American					-0.019	-0.007	-0.019	-0.007	-0.019^{**}	-0.006^{*}	0.001	-0.001	-0.017^{**}	-0.004^{*}
					(0.012)	(0.005)	(0.012)	(0.004)	(0.008)	(0.003)	(0.006)	(0.002)	(0.007)	(0.002)
Male					0.022^{***}	0.007^{***}	0.022^{***}	0.007^{***}	0.021^{***}	0.007^{***}	0.001	-0.000	0.019^{***}	0.005^{***}
					(0.006)	(0.003)	(0.006)	(0.003)	(0.005)	(0.002)	(0.002)	(0.001)	(0.005)	(0.002)
Start Age					-0.002^{***}	-0.000	-0.002^{***}	-0.000	-0.002^{***}	-0.000	0.000	-0.000	-0.002^{***}	-0.000
					(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cohort Size					0.000	-0.000^{*}	0.000	-0.000^{*}	0.000	-0.000^{*}	0.000	-0.000	0.000	-0.000
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	0.249^{***}	-1.064^{***}	-1.389^{***}	0.325^{***}	-1.355^{***}	0.340^{***}	-1.341^{***}	0.334^{***}	-0.663^{***}	0.269^{***}	-0.678^{***}	0.064^{***}	-0.599^{***}	0.190^{***}
	(0.051)	(0.046)	(0.048)	(0.013)	(0.064)	(0.015)	(0.060)	(0.012)	(0.048)	(0.010)	(0.014)	(0.003)	(0.040)	(0.006)
R^2	0.008	0.008	0.019	0.012	0.077	0.048	0.081	0.055	0.082	0.053	0.023	0.017	0.074	0.043
Num. obs.	940	940	940	940	940	940	940	940	940	940	940	940	940	940

Note: Table displays the effect of cohort composition on main sample officer outcomes. The outcome in Column (1) is the average number of arrests of minorities (95 percent of all arrests) in the officer's first 200 shifts as a full (non-probationary) officer; effects estimated using equation (1). The outcomes of Columns (2)-(12) are individual officer fixed effects, estimated using equation (2), for all minority arrests, and minority, Black, and non-Black minority arrests for low-level and serious crimes; effects estimated using equation (3). Columns (13)-(14) replicate Columns (9)-(10) but use shrunken fixed effects, as described in Appendix A.6. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 4: Effect of Cohort Composition on Arrest Propensity - Full Sample

	Shrunken Black Arrest Propensity							
	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort Share Black	-0.211^{**}	0.034^{***}	-0.128^{*}	0.026^{***}				
	(0.096)	(0.012)	(0.069)	(0.006)				
Cohort Share Non-Black Minority	-0.115^{**}	0.008	−0.041	0.005				
	(0.058)	(0.009)	(0.048)	(0.005)				
Cohort Share Minority, Start Age <27					−0.015	0.005	−0.061	−0.010
					(0.052)	(0.009)	(0.093)	(0.025)
Cohort Share Minority, Start Age [27, 32]					−0.090	0.005	−0.128	−0.004
					(0.071)	(0.007)	(0.097)	(0.015)
Cohort Share Minority, Start Age >32					-0.176^{***}	0.021^{***}	-0.189^{***}	0.012
					(0.043)	(0.007)	(0.047)	(0.012)
Controls	Full	Full	Full	Full	Full	Full	Full	Full
Sample	Exams 2006, 2010	Exams 2006, 2010	Full	Full	Full	Full	Exams 2006, 2010	Exams 2006, 2010
R^2	0.232	0.115	0.422	0.154	0.428	0.154	0.235	0.113
Num. obs.	1221	1221	2336	2336	2336	2336	1221	1221

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to arrest Blacks for low-level and serious crimes in even and odd columns, respectively. The propensity is captured by officers' fixed effects using equation (2) and shrunken as described in Appendix A.6. The parameter estimates are based on the specification in equation (3). Full controls refers to the additional controls in Column (13) of Table 3 with exam fixed effects and controls for cohort shares of whites who are start between 27 and 32 and above 32. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5: Effect of Cohort Composition on Arrest Quality Propensity

	Shrunken Black Arrest Propensity							
	Low-Level		Serious		Low-Level		Serious	
	Guilty	Non-Guilty	Guilty	Non-Guilty	Guilty	Non-Guilty	Guilty	Non-Guilty
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort Share Black	−0.010	-0.165^{**}	0.005^{***}	−0.002				
	(0.013)	(0.073)	(0.002)	(0.008)				
Cohort Share Non-Black Minority	−0.003	-0.105^{**}	0.000	−0.003				
	(0.007)	(0.046)	(0.002)	(0.004)				
Cohort Share Minority, Start Age <27					0.005	−0.026	0.000	0.000
					(0.005)	(0.043)	(0.002)	(0.004)
Cohort Share Minority, Start Age [27, 32]					−0.002	−0.091	−0.002	−0.002
					(0.006)	(0.058)	(0.002)	(0.004)
Cohort Share Minority, Start Age >32					-0.009^{**}	-0.154^{***}	−0.002	0.000
					(0.004)	(0.036)	(0.001)	(0.003)
Controls	Full	Full	Full	Full	Full	Full	Full	Full
Sample	Main	Main	Main	Main	Full	Full	Full	Full
R^2	0.068	0.074	0.031	0.036	0.107	0.496	0.033	0.032
Num. obs.	940	940	940	940	2336	2336	2336	2336

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to make high (guilty) and low (non-guilty) quality arrest of Blacks for low-level and serious crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 6: Alternate Samples and Specifications - Main Sample

	Poisson	LPM	First Arresting PO	Crime and Watch Duration Controls	Car Only	FBI Index/Non-Index	Additional Controls	High Crime	Probation	Restricted Sample	First Cohorts
Panel A - Black Low-Level Arrests											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Cohort Share Black	-4.77^{***}	-0.21^{**}	-0.15^{**}	-0.24^{**}	-0.28^{**}	-0.22^{**}	-0.21^{**}	-0.33^{***}	-0.18^{***}	-0.24^{***}	-0.19^{**}
	(1.79)	(0.10)	(0.07)	(0.11)	(0.12)	(0.10)	(0.10)	(0.11)	(0.06)	(0.08)	(0.09)
Cohort Share Non-Black Minority	-2.90^{**}	-0.13^{**}	-0.09^{*}	-0.14^{**}	-0.18^{**}	-0.14^{**}	-0.19^{**}	-0.12^{*}	-0.11^{**}	-0.15^{***}	-0.25^{***}
	(1.19)	(0.06)	(0.04)	(0.07)	(0.08)	(0.06)	(0.08)	(0.06)	(0.04)	(0.05)	(0.03)
R^2	0.14	0.09	0.07	0.08	0.08	0.08	0.09	0.09	0.09	0.10	0.09
Panel B - Black Serious Arrests											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Cohort Share Black	3.48^{***}	0.06^{***}	0.05^{***}	0.06^{***}	0.08^{***}	0.07^{***}	0.04^{**}	0.07^{*}	0.10^{***}	0.14^{***}	-0.00
	(0.85)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.03)	(0.02)
Cohort Share Non-Black Minority	1.87^{***}	0.02^{*}	0.02	0.02	0.03^{*}	0.04^{***}	0.00	0.04	0.03	0.03	-0.01
	(0.59)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Controls	Full	Full	Full	Full	Full	Full	Additional	Full	FTO Demos	Full	No Cohort Size
R^2	0.13	0.05	0.05	0.05	0.05	0.08	0.05	0.06	0.03	0.06	0.03
Num. obs.	890	940	940	940	940	940	940	509	940	908	385

Note: Table displays the effect of cohort diversity on officer propensity to arrest Blacks for low-level (Panel A) and serious crimes (Panel B) for main sample officers. The propensity is captured by officers' fixed effects using equation (2), Column (6) uses shrunken fixed effects. The parameter estimates are based on the specification in equation (3) (unless otherwise specified), with controls denoted as Full referring to the specification in Column (13) of Table 3, and additional or removed controls denoted in the table. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level (unless otherwise specified) are in parentheses.

Columns (1),(2): Results from estimating officer fixed effects using equation (4) and modifying equation (2) with the dependent variable (arrests of type k) being whether (1) or not (0) the officer made at least one arrest of type k during their shift. Column (3): The fixed effects used as dependent variables were recovered from estimating equation (2) with an arrest only contributing toward the dependent variable if the officer was the first arresting officer.

Column (4): The fixed effects used as dependent variables were recovered from estimating equation (2) including crime (violent, property, sex, drug, domestic, and other) and watch duration second degree polynomials in the estimation.

Column (5): The fixed effects from estimating equation (2) on the sample of shifts where officers were in cars (>85Column (6): Reclassifies serious and low-level arrests as index and non-index based on the FBI UCR classification and excludes warrant arrests due to unknown crime types. Column (7): Uses shrunken fixed effects (e.g., used in Columns (13)-(14) of Table 3) with additional controls for officer being in the military, speaking Spanish, having a Bachelors degree or higher and cohort shares of females, military, Spanish speakers, and Bachelor degrees or higher, and cohort mean start age. Column (8): Results use new officers from the main sample whose average district of assignment was above the 75th percentile of violent index crime in Chicago. Violent index crime rates are violent index crimes (murder, rape, robbery, aggravated assault) in a month, based on Chicago City Data Portal crime data, per 10,000 population, based on 2010 Census estimates.

Column (9): Results are from estimating equation (3) on fixed effects of new officers recovered during their probationary periods only, with additional controls for share of Field Training Officers (FTOs) that were Black, Hispanic, and other (non-white) race.

Column (10): Results from estimating the first stage (equation (2)) on a restricted sample of officer assignments, only including assignments in 2012 to 2015, with no additional information codes, and only regular watch assignments (the three main shifts), using assignment fixed effects as described in Ba et al. (2021).

Column (11): The sample is the subset of the main sample of recruits who started within 5 months of the first 2012 cohort, which is 7 cohorts. Fixed effects are shrunken. Due to few cohorts, Webb (2014) standard errors are used clustered at the cohort level, and I do not control for cohort size.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 7: Alternate Samples and Specifications - Full Sample

	Poisson	LPM	First Arresting PO	Crime and Watch Duration Controls	Car Only	FBI Index/Non-Index	Additional Controls
	Panel A - Black Low-Level Arrests						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cohort Share Minority, Start Age <27	−1.04	−0.03	−0.03	−0.04	−0.05	−0.03	−0.04
	(1.37)	(0.06)	(0.04)	(0.07)	(0.08)	(0.06)	(0.06)
Cohort Share Minority, Start Age [27, 32]	-2.84^{*}	−0.10	−0.09	−0.12	−0.15	−0.09	−0.10
	(1.48)	(0.08)	(0.05)	(0.09)	(0.10)	(0.08)	(0.07)
Cohort Share Minority, Start Age >32	-5.43^{***}	-0.21^{***}	-0.14^{***}	-0.23^{***}	-0.29^{***}	-0.21^{***}	-0.23^{***}
	(1.11)	(0.05)	(0.03)	(0.05)	(0.06)	(0.05)	(0.07)
R^2	0.73	0.53	0.43	0.50	0.58	0.49	0.44
	Panel B - Black Serious Arrests						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cohort Share Minority, Start Age <27	1.30	0.03	0.02	0.03	0.03	0.03	0.01
	(0.97)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Cohort Share Minority, Start Age [27, 32]	1.27	0.04^{**}	0.03^{*}	0.03^{*}	0.04^{**}	0.04^{**}	0.01
	(0.98)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Cohort Share Minority, Start Age >32	3.91^{***}	0.07^{***}	0.06^{***}	0.07^{***}	0.09^{***}	0.09^{***}	0.03^{***}
	(0.83)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Controls	Full	Full	Full	Full	Full	Full	Additional
R^2	0.72	0.41	0.44	0.37	0.46	0.71	0.16
Num. obs.	2207	2336	2336	2336	2336	2336	2336

Note: Table displays the effect of cohort diversity on officer propensity to arrest Blacks for low-level (Panel A) and serious crimes (Panel B) for main sample officers. The propensity is captured by officers' fixed effects using equation (2), Column (6) uses shrunken fixed effects. The parameter estimates are based on the specification in equation (3) (unless otherwise specified), with controls denoted as Full referring to the specification in Column (13) of Table 3 and controls for cohort shares of whites who are start between 27 and 32 and above 32, and additional or removed controls denoted in the table. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level (unless otherwise specified) are in parentheses.

Columns (1),(2): Results from estimating officer fixed effects using equation (4) and modifying equation (2) with the dependent variable (arrests of type k) being whether (1) or not (0) the officer made at least one arrest of type k during their shift. Column (3): The fixed effects used as dependent variables were recovered from estimating equation (2) with an arrest only contributing toward the dependent variable if the officer was the first arresting officer.

Column (4): The fixed effects used as dependent variables were recovered from estimating equation (2) including crime (violent, property, sex, drug, domestic, and other) and watch duration second degree polynomials in the estimation.

Column (5): The fixed effects from estimating equation (2) on the sample of shifts where officers were in cars (>85Column (6): Reclassifies serious and low-level arrests as index and non-index based on the FBI UCR classification and excludes warrant arrests due to unknown crime types. Column (7): Uses shrunken fixed effects (e.g., used in Columns (13)-(14) of Table 3) with additional controls for officer being in the military, speaking Spanish, having a Bachelors degree or higher and cohort shares of females, military, Spanish speakers, and Bachelor degrees or higher.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 8: Interaction of Cohort Diversity and Officer Race

	Shrunken Low-Level Black Arrest Propensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	-0.121					
	(0.119)					
Cohort Share Non-Black Minority	-0.043					
	(0.053)					
White x Cohort Share Black	-0.150					
	(0.110)					
White x Cohort Share Non-Black Minority	-0.144^{**}					
	(0.059)					
Cohort Share Minority, Start Age <27		0.006			-0.012	-0.013
		(0.052)			(0.051)	(0.051)
Cohort Share Minority, Start Age [27, 32]		-0.038			-0.087	-0.088
		(0.067)			(0.067)	(0.067)
Cohort Share Minority, Start Age >32		-0.166^{***}			-0.165^{***}	-0.166^{***}
		(0.050)			(0.041)	(0.040)
White x Cohort Share Minority, Start Age <27		-0.052				
		(0.038)				
White x Cohort Share Minority, Start Age [27, 32]		-0.113^{**}				
		(0.050)				
White x Cohort Share Minority, Start Age >32		-0.021				
		(0.053)				
Cohort Share Black, Start Age <=32			-0.091	-0.083		
			(0.120)	(0.110)		
Cohort Share Black, Start Age >32			-0.231^{**}	-0.207^{**}		
			(0.113)	(0.092)		
Cohort Share Non-Black Minority, Start Age <=32			-0.034	0.012		
			(0.050)	(0.043)		
Cohort Share Non-Black Minority, Start Age >32			-0.142^{***}	-0.116^{**}		
			(0.045)	(0.049)		
White x Cohort Share Black, Start Age <=32				-0.028		
				(0.087)		
White x Cohort Share Black, Start Age >32				-0.058		
				(0.131)		
White x Cohort Share Non-Black Minority, Start Age <=32				-0.102^{**}		
				(0.042)		
White x Cohort Share Non-Black Minority, Start Age >32				-0.053		
				(0.050)		
Cohort Share Minority and Female					-0.140^{**}	-0.114^{*}
					(0.064)	(0.061)
Cohort Share White and Female					-0.049	-0.001
					(0.093)	(0.100)
White x Cohort Share Minority, Female						-0.052
						(0.052)
White x Cohort Share White, Female						-0.094
						(0.081)
Cohort Share White, Start Age >=27		-0.101^{***}	-0.064	-0.082^{**}	-0.067^{*}	-0.067^{*}
		(0.034)	(0.046)	(0.040)	(0.039)	(0.039)
White x Cohort Share White, Start Age >=27		0.043		0.030		
		(0.027)		(0.034)		
Controls	Full	Full	Full	Full	Full	Full
Sample	Main	Full	Full	Full	Full	Full
R^2	0.079	0.430	0.428	0.430	0.433	0.434
Num. obs.	940	2336	2336	2336	2336	2336

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to arrest Blacks for low-level and serious crimes. The propensity is captured by officers' fixed effects using equation (2) and shrunken as described in Appendix A.6. The parameter estimates are based on the specification in equation (3) with the addition of terms for cohort shares interacted with an officer being white. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

A Appendix A

A.1 Entrance into the CPD Police Academy

In order to become an officer in the CPD, applicants must first meet multiple qualifications before applying to take the entrance exam. For example, by the time of starting at the academy, one must be a US citizen, a resident of Chicago, have sufficient credit hours at a college or university, and meet the age requirement (Pritchard (2013)). Potential applicants meeting these qualifications can apply to take the CPD entrance exam, and they will be notified of the test date and location after the application period ends (CPD (2017)).[46]

Applicants who pass the written exam are then assigned a random lottery number indicating the order in which they will be called into the academy. Random assignment to the academy was not always the case; it was introduced in the early 1990's in an attempt to increase diversity (Kass and Blau (1991)). After an applicant's number is drawn, they must pass a background check, drug screening, and medical, psychological and physical exams (Pritchard (2013)). Upon passing these requirements, potential officers are admitted into the academy.

There are usually tests once every 2 or 3 years (not including makeup exams)—but in 2006 there were four exams issued.[47] Generally, thousands of people take the CPD's written exam and a large portion of them meet the minimum passing score (see Figure B.1). Given the large number of passing applicants, many do not ever have their numbers called before the applicant list is retired. Despite my best efforts, I have not been able to obtain any indication of when the applicant lists are retired (according to the CPD such documentation may not even exist). Also, applicants from a test are likely to be admitted possibly years after they took the test initially, and their entrance into the academy likely occurs while more applicants are taking a new test. This makes identifying which cohorts come from which tests (i.e. the pool from which officers are randomly assigned) difficult.

To the best of my knowledge, the main sample (July 2012 to May 2014) cohorts are an exception, and these cohorts all came from the same exam issued in December of 2010 (see Figure B.1). The December 2010 exam was the last exam issued before the December 2013 exam. The only sizable cohort to enter in 2011 was on October 17, 2011, then about 8 months pass until the first sizable cohort of 2012 started on July 02, 2012. Following this,

[46]As late as the 2013 exam, veterans began to received preference in their lottery numbers– though this is not well defined in documentation. However, this preference is unlikely to be important considering almost all (over 95%) of recruits have military experience in the full sample. This very large amount of veterans is consistent with more recent estimates from the Office of the Inspector General (Ferguson and Witzburg (2021)).

[47]One is labeled a '2005' exam in Figure B.1, but it took place in February 2006.

there were a total of 7 sizable cohorts starting between July and December of 2012. Then, there is continuous intake of cohorts until May of 2014, when there is a 3 month gap until the next cohort. Given that it takes time for the CPD to draw in passing recruits and give them their multiple examinations, I believe that the main sample cohorts were all drawn from the December 2010 exam.

Further supporting this is the change in the composition of cohorts before and after 2012. As shown in Panel A of Figure A.1, the 2011 cohort has a higher share Black than almost every cohort in the 2012-2014 period, while it is within the range of the Exam 2006 cohorts (likely drawn from the 2006 tests). Similar patterns emerge when looking at share of the cohort which speaks Spanish (see Panel B of Figure A.1), where all of the 2006 cohorts have strictly smaller shares of Spanish speakers compared with any 2010 cohort. Finally, minimum start age (Panel C) increases successively for each of the pre-2010 cohorts (as expected since these recruits have been waiting at least 4 years to enter), while it decreases slightly in the first 2010 cohort and significantly in the second 2010 cohort. Anecdotally, an officer I spoke with who started the academy in 2012 confirmed that their cohort was comprised of 2010 test takers.

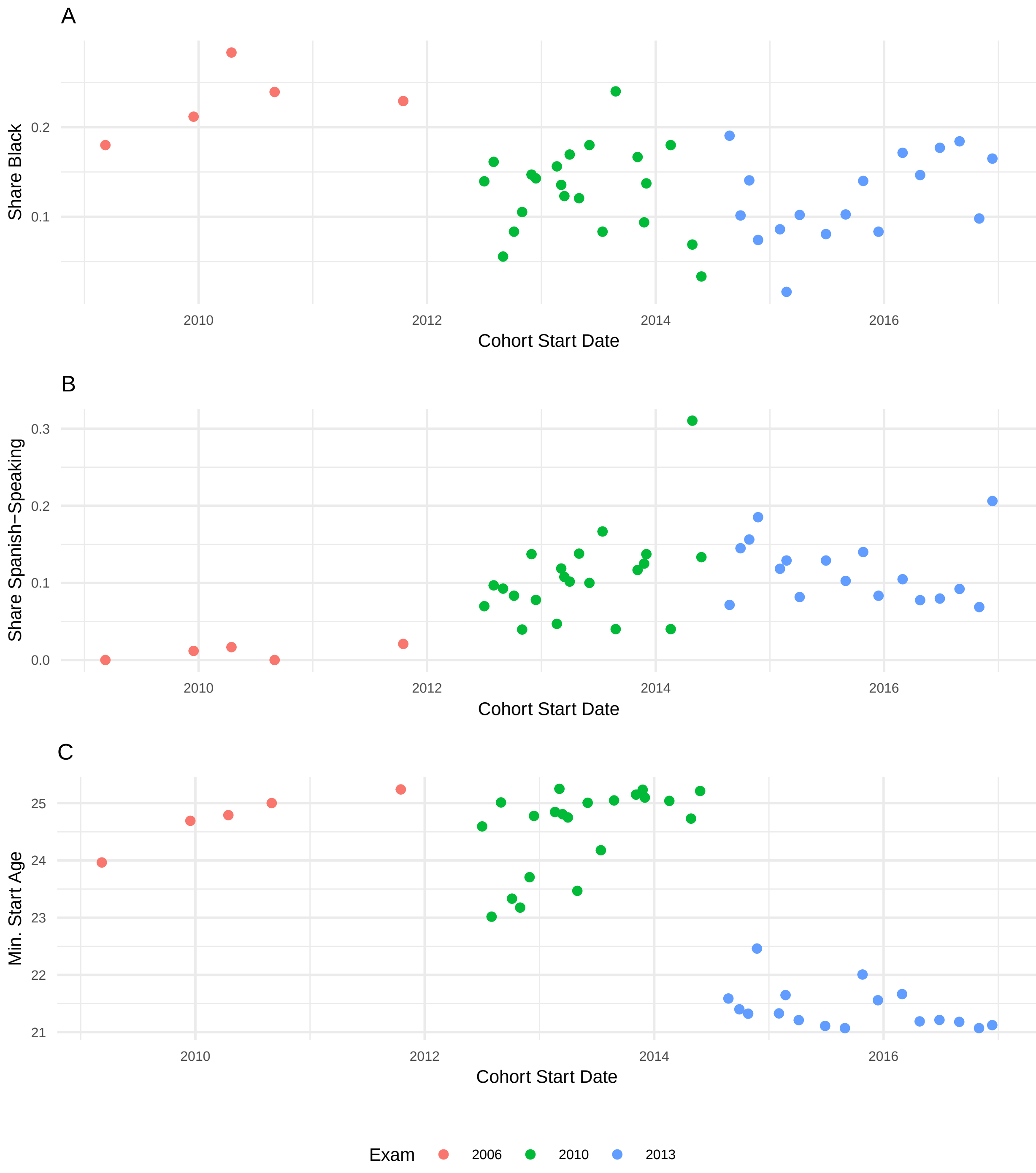
In separating the Exam 2006 cohorts (starting in 2009 and ending in 2011) from the Exam 2010 cohorts, and determining if all Exam 2006 cohorts actually came from the four 2006 exams (and not the 2004 exam), I use posts on a police forum (https://forum.officer.com/) in 2009, 2010, and 2011. One poster on November 17th, 2009, states: "Just got the call... the academy starts December 16th... My number is 1036, and I am a June 06 tester." (Chicago_mwk (2010), pg.29). December 16th, 2009, is the start date of the first cohort in my full sample. This is followed by a flurry of other posters stating their numbers also got called for the same start date. The only cohort before it was in March of 2009, which according to a poster in on March 6th 2009, "From what I know [the March 2006 cohort] it's a mix of Feb 06 and early June 06 testers." (Chicago_mwk (2010), pg.9). Overall, this indicates the 2009 and 2010 cohorts came from Exam 2006 test takers only.

Next, the main question is did the single 2011 (in October) cohort end the Exam 2006 cohorts or start the Exam 2010 cohorts? According to a different thread on the same site, a poster on December 4th, 2010, states: "With roughly 40 candidates ready for hire off the 2006 test, and a new test next week, its about time we started this thread. For those who are wondering, the last of the 2006 list (40 people) were scheduled to start on 01 November [2010] but according to my BI who I call twice every month, the class has been pushed back and only the fine folks at city hall know the date. In my humble opinion city hall is waiting on the new year [2011] to start our class because of the new budget and the new pension system for new hires." (neverlose357 (2010), pg.1) On September 30th, 2011, a poster states that their cohort ("2011-1") will "soon fill the halls of the Chicago Police Academy" (neverlose357 (2010), pg. 6), and another poster, on October 18th, 2011, (one day after the 2011 cohort starts in the data) states that the class has "About 50" recruits (49 in the data). The rest of this forum discusses the composition of this cohort. It is stated that this cohort will exhaust the rest of the 2006 applicants (at least 32) and fill the rest either with 2006 applicants who won appeals or 2010 testers. So, based on these discussions, the single 2011 cohort finished off the 2006 Exam cohorts, and was potentially mixed with a small number of 2010 Exam takers– though this seems to be an unusual practice and only a result of the small number of potential recruits in the 2006 tests (neverlose357 (2010), pg. 3). Because of this issue with mixing a single cohort, Table B.6 displays the results of Table 4 excluding the 2011 cohort, and the results are highly similar.

After May 2014, the cohorts until December 2016 (the last cohort I use in the extended sample) are from the 2013 test. The 2013 test recruits had the new feature that they were permitted to begin the academy at the age of 21, lower than the previous requirement of 23 (Pritchard (2013)). As can be seen in Panel C of Figure A.1, the lowest starting age per cohort drops to 21 after the May 2014 cohort. Thus, I can distinguish between the 2010 and 2013 test cohorts using this feature. The end of the 2013 test cohorts occurs after the final cohort in the full sample in December of 2016. Even though there was a test issued in April 2016, based on forum posts about 2016 recruitment the 2016 test-takers had not begun to be drawn in by the end of 2016. Following many 2016 test takers wondering when their cohorts would be drawn in, one poster stated on December 26, 2016, "People that took the exam in 2013 are still being processed. I believe about 9k people passed the written exam this year" (Aendos (2015), pg. 138). So, I am confident that the Exam 2013 sample does not contain 2016 test cohorts. Based on the panels in Figure A.1, there is fairly consistent cohort composition across the Exam 2013 cohorts. While extending my cohorts beyond December 2016 is possible, because my panel data extends to 2018 (overlapping with court data and outcomes), including the first cohorts in 2017 would not contribute much to my analysis as these officers would have less than 6 months of observations in the panel data after their probationary period.

As I am less confident as I move away from the Exam 2010 cohorts – not knowing the beginning of the Exam 2006 cohorts and being restricted in the panel with post-2016 cohorts – I focus on them (the main sample) in my analysis. However, as incorporating the Exam 2006 and Exam 2013 cohorts into my analysis provides results which are generally consistent with my main sample results and they provide a significant increase in sample size, I use them collectively as well.

Figure A.1: Composition of Cohorts by Start Date



Note: Figure displays the share of cohorts with more than 10 starting members that are Black (Panel A) and speak Spanish (Panel B), and the lowest starting age (Panel C) by the cohort start date, from 2009 to 2016. Exam denotes the time period during which the cohorts started and assumes cohorts in the same period were in the same test pool.

A.2 Random Assignment

Given that the timing of when a recruit can enter the academy is determined by a random lottery number, cohorts are as-good-as-randomly assigned, and I provide empirical evidence for this by testing for violations of the random assignment assumption. Table A.1 displays the p-value of a joint F-test resulting from a multinomial logit of assigned cohort on officer characteristics for each of the three exams separately. The main sample (Exam 2010) cohorts have the highest p-values and are far from statistically significant even when including additional officer factors such as education, military status, and Spanish language ability. The Exam 2013 cohorts have a non-significant p-value with the main controls (race, gender, start age); however, it becomes marginally significant when additional controls are introduced.[48] There are only 5 Exam 2006 cohorts, yet the p-value is not statistically significant with the limited controls– adding additional controls produces a statistically significant p-value, however there are more predictors than potential cohorts. These results indicate that officer characteristics are not predictive of assigned cohort, particularly in the main sample.

Table A.1: Multinomial Logit for Cohort Assignment

Main Sample	Exam	Controls	Multinomial Logit P-Value	N Recruits	N Cohorts
No	2006	Minority, Gender, Start Age	0.155	360	5
No	2006	+ Military, Spanish, High Edu.	0.000	360	5
Yes	2010	Minority, Gender, Start Age	0.860	1135	21
Yes	2010	+ Military, Spanish, High Edu.	0.444	1135	21
No	2013	Minority, Gender, Start Age	0.307	1253	17
No	2013	+ Military, Spanish, High Edu.	0.081	1253	17

Note: Table reports the p-value of the joint F-test on the coefficients of a multinomial logit regressing assigned cohort on officer characteristics for each exam period for two sets of controls. The limited controls include the officer being a minority, start age, and gender; the second set of controls adds if they were in the military, if they speak Spanish, and if they have a Bachelors degree or higher.

For additional evidence of random assignment, I test if officer characteristics are significantly associated with cohort composition. Table A.2 displays the results of regressing an officer's cohort characteristics on their individual characteristics. Columns (1)-(3) focus on the main sample with baseline controls (minority, gender, start age); Columns (4)-(6) also focus on the main sample but include additional variables (military status, Spanish ability, and education level); Columns (7)-(9) repeat the analysis of Column (4)-(6) with the full sample. Based on Columns (1)-(3), the baseline controls explain very little of the variation in cohort composition and all statistically significant coefficients are economically insignificant: being Male has the largest statistically significant effect, but it implies that being

[48]The CPD's demographic data often combines race and ethnicity into a single variable. For expositional purposes and due to the data used, I will refer to 'Hispanic' as a distinct racial group.

male is associated with only a 0.21% increase in average cohort age relative to the mean. As expected, being male is negatively associated with cohort share male, just as being a minority is negatively associated with cohort share minority– since cohort shares exclude the officer in question, it reduces the pool of officers with that characteristic, as noted in Guryan, Kroft, and Notowidigdo (2009). The additional officer characteristics in Columns (4)-(6) are also generally statistically and economically insignificant. Using the full sample in Columns (7)-(9) supports the lack of economic or statistical significance in the full sample as well.

Given that cohorts begin successively and not all at the same time, it is likely some amount of selection out of the academy by officers who give up, find other jobs, are no longer eligible (too old, moved out of Chicago, could not pass the physical exam, etc.)– though this may be limited in Chicago relative to other police departments as the CPD is highly oversubscribed and a well-paying department. We want to ensure that this delayed entrance and selection does not significantly alter the composition of recruits. Column (10) of Table A.2 regresses when the officer started at the academy (in years since 2009) on officer characteristics for the full sample. An officer being a minority, in the military, a Spanish-speaker, or highly educated (Bachelors or higher) have no statistically significant effect on when they start at the academy. Unsurprisingly, officer start age is statistically significant and positively associated with start date, but as the coefficient is less than 1 it implies there is censoring (at 40) and likely some selection out for aging officers. A recruit being male is associated with an earlier start date, implying male applicants may exit the pool quicker than female applicants, though being male is only associated with starting the academy about 0.055 years before female applicants, which equates to about 20 days. While attrition from the sample pool is almost certain, the evidence presented indicates it is not likely to significantly impact the composition of cohorts or be associated with differences in officer unobservables. Furthermore, applicants wait over a year before the first cohort is called in, meaning the least committed applicants likely select out once they receive their numbers and understand where they are in the pool.

Table A.2: Balance Regressions

	Main Sample						Full Sample			
	Cohort Mean Age	Cohort Share Minority	Cohort Share Male	Cohort Mean Age	Cohort Share Minority	Cohort Share Male	Cohort Mean Age	Cohort Share Minority	Cohort Share Male	Start Date (years)
Minority	0.014	0.002	−0.005	−0.003	0.001	−0.008**	0.022	0.000	−0.005**	0.028
	(0.023)	(0.002)	(0.003)	(0.021)	(0.002)	(0.003)	(0.018)	(0.001)	(0.002)	(0.022)
Male	0.064**	−0.009***	0.005	0.063**	−0.009***	0.005	0.001	−0.002	0.001	−0.055**
	(0.028)	(0.003)	(0.004)	(0.027)	(0.003)	(0.003)	(0.020)	(0.004)	(0.002)	(0.027)
Start Age	0.000	0.000*	0.000	0.000	0.001*	0.000	0.002	−0.000	0.001*	0.010***
	(0.004)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.003)
Military				−0.078	0.003	−0.005	0.097	−0.003	0.003	0.194
				(0.060)	(0.005)	(0.009)	(0.097)	(0.010)	(0.007)	(0.169)
spanish				0.076	0.005	0.012**	0.013	0.004	0.007	−0.003
				(0.058)	(0.005)	(0.005)	(0.048)	(0.003)	(0.005)	(0.062)
High Edu				−0.028	0.001	−0.004	−0.039	0.004*	−0.003	−0.024
				(0.040)	(0.002)	(0.003)	(0.025)	(0.002)	(0.003)	(0.032)
Exam 2010							0.489**	0.022	−0.043***	2.875***
							(0.245)	(0.034)	(0.015)	(0.293)
Exam 2013							−1.420***	−0.010	−0.021	5.496***
							(0.255)	(0.035)	(0.018)	(0.331)
Intercept	29.976***	0.792***	0.482***	30.067***	0.788***	0.489***	29.412***	0.781***	0.521***	0.945***
	(0.208)	(0.016)	(0.019)	(0.203)	(0.017)	(0.021)	(0.218)	(0.039)	(0.015)	(0.349)
R^2	0.002	0.014	0.003	0.006	0.016	0.009	0.725	0.093	0.062	0.894
Adj. R^2	−0.000	0.012	0.001	0.001	0.011	0.003	0.724	0.090	0.060	0.893
Num. obs.	1135	1135	1135	1135	1135	1135	2748	2748	2748	2748

Note: Table display results for balance regression tests. Each column displays the coefficients of officer characteristics on their cohort composition (Columns (1)-(9)) and start date (Column (10)) for main sample (Columns (1)-(6)) and full sample (Columns (7)-(10)) officers. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level (unless otherwise specified) are in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Finally, there are two points relating to random assignment and identification that should be made. First, identification of peer composition's effect on outcomes is possible thanks to variation in cohort composition differing due to random draws. If there were, for example, only two large cohorts from an exam, the variation is expected to be small. In the main sample, there are 21 cohorts, and as expected, the variation in cohort composition is larger in cohorts that are smaller: the correlation between cohort size and the absolute difference between cohort share minority and the mean cohort share minority is -0.4. Second, there may be a concern that cohort size is related to unobservable trends or departmental demands that influence officer outcomes. For example, the department may be expecting a low-crime period, and thus draw in fewer officers (a smaller cohort), which may be an issue for estimates relying on cohort composition variation. However, this is unlikely because the CPD must draw in applicants more than 1.5 years before they can actually begin as police officers, and these decisions take into account not only (likely imprecise) crime projections but also city budgets, staffing demands, and various other concerns.

A.3 Attrition

If the likelihood of attrition from the sample is impacted by diversity of one's cohort, then results in my estimation may be driven by selection bias rather than actual peer effects. In Table A.3, I present results for logistic regressions where each outcome is a form of attrition for officers in the main sample cohorts (Columns (1)-(5)) and full sample (Column (6)-(10)).

The outcome of Column (1) and (6) is whether the officer is not in the daily assignment (AA) data (52 main sample recruits). The outcome of Column (2) and (7) is outcome is whether the officer, conditional on being in the AA data, spent too much or too little time in the academy or probationary period (113 main sample recruits). The outcome of Column (3) and (8) is is whether the recruit was not in the final AA data, conditional on the previous two restrictions, meaning they were not matched to the salary and rank data as a police officer (24 main sample recruits). The outcome of Column (4) and (9) is any form of attrition across all recruits, including whether fixed effects could be recovered. The outcome of Columns (5) and (10) is any attrition from the sample (as in (4) and (9)), and also does not appear in the training data (data on specific courses the recruits attended). As displayed across all columns for the main sample, there is no statistically significant predictor of any form of attrition with respect to cohort composition (neither cohort diversity nor mean age), thus it is unlikely that attrition driven selection is driving my results. For the full sample, only peer age is ever statistically significant (Column (8)).

Another form of attrition is sample attrition after the recruits exit the academy, become full officers, and are present in the assignment data, e.g. cohort diversity being related to when officers choose to retire or exit the assignment data. While this may cause some officers to be more represented in the sample than others, the fixed effects recovered for the main sample are based on over 100 observations for almost all officers (96.6% of main sample recruits). I test for sample-exiting attrition in Table A.4. Column (1) study the relationship between cohort share minority and officer's number of observations in the assignment data used to estimate fixed effects, showing no significant relationship. Column (2) shows that cohort diversity has no effect on whether the officer exists in the salary and unit history data (which contains officers not in the assignment data) at the end of 2018. Column (3) shows cohort diversity has no significant effect on likelihood of being promoted by the end of 2018 and that average peer age has a statistically significant but economically small effect. Column (4) shows the likelihood of being in a non-geographic unit at the end of 2018 is not significant impacted by cohort composition.

Table A.3: Attrition from Sample

	Main Sample					Full Sample				
	Not in AA	Training Time Violation	Not in Final AA	Any Attrition	Not in Trainings	Not in AA	Training Time Violation	Not in Final AA	Any Attrition	Not in Trainings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cohort Share Minority	−1.06	1.31	0.71	0.22	−1.48	0.19	1.46	−0.22	0.22	0.01
	(3.14)	(3.81)	(3.49)	(2.77)	(3.00)	(1.95)	(3.06)	(2.60)	(2.73)	(2.12)
Cohort Mean Age	0.51	−0.03	−0.72	0.05	−0.18	0.09	−0.02	-0.63^{**}	0.05	−0.32
	(0.37)	(0.29)	(0.41)	(0.23)	(0.32)	(0.18)	(0.29)	(0.22)	(0.23)	(0.23)
Black	−0.24	0.19	-15.83^{***}	0.03	0.26	−0.06	0.24	−0.14	0.03	0.25
	(0.47)	(0.41)	(0.40)	(0.34)	(0.30)	(0.22)	(0.23)	(0.43)	(0.33)	(0.15)
Hispanic	0.43	0.19	−0.03	0.25	0.28^{*}	0.30	0.25	0.10	0.25	0.25^{**}
	(0.36)	(0.24)	(0.55)	(0.19)	(0.14)	(0.20)	(0.15)	(0.33)	(0.19)	(0.09)
Asian/Native American	−0.68	0.47	-16.04^{***}	−0.01	0.35	−0.82	0.07	−0.86	−0.01	−0.10
	(0.90)	(0.45)	(0.43)	(0.40)	(0.32)	(0.70)	(0.37)	(0.97)	(0.39)	(0.23)
Male	−0.05	−0.40	0.29	−0.28	−0.19	−0.21	−0.13	0.00	−0.28	−0.14
	(0.39)	(0.24)	(0.59)	(0.23)	(0.19)	(0.25)	(0.18)	(0.34)	(0.23)	(0.12)
Start Age	0.02	−0.02	-0.13^{*}	−0.02	-0.04^{*}	−0.00	−0.01	-0.10^{**}	−0.02	-0.03^{*}
	(0.02)	(0.03)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.01)
Cohort Size	0.00	−0.03	−0.00	−0.02	−0.01	−0.00	-0.02^{*}	−0.00	−0.02	0.00
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Exam 2010						−0.43	−0.64	0.33	-22.64^{***}	−0.66
						(0.37)	(0.35)	(0.48)	(0.62)	(0.40)
Exam 2013						−0.38	-1.22^{**}	-1.21^{*}	0.02	-2.26^{***}
						(0.44)	(0.41)	(0.56)	(0.58)	(0.56)
Intercept	−18.97	0.58	21.20	−1.45	6.66	−5.17	−0.41	17.91^{**}	21.19^{***}	10.17
	(11.20)	(7.68)	(13.34)	(6.07)	(10.13)	(5.41)	(7.30)	(6.86)	(5.91)	(6.60)
AIC	433.52	710.84	226.46	1039.10	1347.12	1009.95	1439.73	538.15	1043.10	2826.49
Log Likelihood	−207.76	−346.42	−104.23	−510.55	−664.56	−493.97	−708.86	−258.07	−510.55	−1402.25
Num. obs.	1135	1083	970	1135	1135	2748	2626	2409	2748	2748

Note: Table display the logistic regression estimates of cohort and officer observables on officer attrition for various reasons from the main sample (July 2012 - May 2014 cohorts). The dependent variables for Columns (1)-(3) are: (1) whether or not the officer is not matched in the assignment data; (2) whether or not the officer is dropped due to spending too much or too little time in the academy or probationary period; (3) whether or not the officer is not in the final assignment data, meaning they were matched in the salary and unit history data and spent some time as a D1 officer in units 1-25; (4) attrition for any of the listed reasons or if no fixed effects could be recovered due to too few observations. Columns (2)-(3) are estimated on the sample of recruits which were not dropped due to the previous column's reason, and Column (4) is estimated on all initial sample recruits. Column (5) is whether or not the officer appears in the training cohort sample, meaning no attrition (Column (4)) and is in the data on trainings/classes. Columns (6)-(10) replicate (1)-(5) for the full sample. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.4: Attrition out of Sample

	N. Obs in Data	Exit Data	Promoted at End	Specialized Unit at End
	(1)	(2)	(3)	(4)
Cohort Share Minority	-315.73	-0.03	-0.06	-0.07
	(256.54)	(0.13)	(0.12)	(0.11)
Cohort Mean Age	-52.42^{*}	0.02	-0.03^{*}	-0.02
	(24.14)	(0.01)	(0.01)	(0.01)
Black	-28.94	-0.01	-0.04^{*}	-0.07^{***}
	(15.56)	(0.02)	(0.02)	(0.02)
Hispanic	16.39	0.01	-0.03	-0.05^{***}
	(12.10)	(0.01)	(0.02)	(0.01)
Asian/Native American	24.21	0.03	-0.04	-0.01
	(23.06)	(0.03)	(0.03)	(0.03)
Male	99.24^{***}	0.02	0.01	0.01
	(20.26)	(0.02)	(0.01)	(0.01)
Start Age	2.03	0.00^{**}	-0.00	-0.00^{**}
	(1.40)	(0.00)	(0.00)	(0.00)
Cohort Size	-0.77	0.00	0.00	0.00^{*}
	(0.52)	(0.00)	(0.00)	(0.00)
Exam 2010	-287.01^{***}	0.06^{*}	-0.04	-0.04^{**}
	(43.96)	(0.02)	(0.02)	(0.02)
Exam 2013	-634.17^{***}	0.16^{***}	-0.26^{***}	-0.21^{***}
	(51.18)	(0.03)	(0.02)	(0.02)
Intercept	2461.07^{***}	0.11	1.04^{**}	0.87^{**}
	(658.18)	(0.34)	(0.34)	(0.29)
Mean Dep. Var	446.17	0.89	0.11	0.07
R^2	0.42	0.03	0.09	0.09
Adj. R^2	0.42	0.02	0.08	0.08
Num. obs.	2497	2748	2441	2441

Note: Table display the linear regression estimates of cohort and officer observables on officer observations and other measures of attrition for the full sample. The dependent variables are the officer's number of observations (shifts) used to estimate fixed effects in the daily panel data (Column (1)), whether or not the officer is in the salary and unit history data which contains non-D1 officers and units outside of the assignment data (Column (2)), whether the officer has been promoted by the end of 2018 (Column (3)), whether the officer is in a specialized unit at the end of 2018 (Column (4)). Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.4 Confounding Assignments

While cohort composition is not related to a new officer's likelihood of attrition, cohort composition may influence where an officer works. The CPD's unit (district) assignment process is seniority based, with new recruits being assigned based on departmental demand, avoiding very long commutes for officers.[49] and possibly a desire to have officers reflect the communities they police. As a result, it is possible that assignment is influenced by one's cohort's composition, since if Black officers are more likely to be initially placed in Black areas (for example, if the department attempts to place Black officers in districts with Black civilians or those officers are more likely to live close by), having more Black cohort-mates may reduce another officer's probability of being placed in a majority-Black district. In this subsection, I explore the effect of peer diversity on assignment.

How departmental need and initial assignments are determined is not clear, and there does not seem to be a significant correlation between a recent decline in the number of officers in a unit and the share of new recruits serving in that unit. In the data, however, there is a clear relationship between the race of officers serving in a district and that of the district's population. Furthermore, as officers gain seniority and the ability to leave high-crime (high minority population) districts, they do so, meaning the districts in need of officers are higher crime areas.

Table A.5 displays the characteristics of the average unit (district) in which officers work for officers in the full panel. These simple regressions explain some amount of variation in officer assignments. Based on the results, new officers are much more likely to be placed in high crime areas both during and after their probationary periods—this may be partially explained by those units demanding the most officers and higher seniority officers transferring to less dangerous areas. As expected, assignments are influenced by officer race: being Black increases the share of Black civilians in the average district in which an officer works by over 8 percentage points. Similarly, there is a clear relationship between officers being Black and their districts having higher crime and lower income. These results are, however, for all officers in the panel which means the estimates are noisy due to significant changes in crime and recruit composition over time.

[49] This information is based on conversations with a retired officer.

Table A.5: Characteristics of Average Working District

	Violent Crime	Median Income	Share Black Pop.	Share Hispanic Pop.
	(1)	(2)	(3)	(4)
Probationary	2.60^{***}	-8169.67^{***}	0.15^{***}	-0.02^{***}
	(0.12)	(304.27)	(0.01)	(0.00)
Recruit	4.70^{***}	-11446.91^{***}	0.23^{***}	-0.04^{***}
	(0.16)	(358.97)	(0.01)	(0.01)
Female	-0.20	560.99	-0.01	-0.01
	(0.12)	(345.19)	(0.01)	(0.00)
Black	3.26^{***}	-5486.64^{***}	0.26^{***}	-0.13^{***}
	(0.14)	(404.67)	(0.01)	(0.00)
Hispanic	0.61^{***}	-1814.42^{***}	0.02^{**}	0.02^{***}
	(0.13)	(333.13)	(0.01)	(0.00)
Other Race	-0.52^{*}	1156.56	-0.04^{***}	-0.01
	(0.27)	(720.41)	(0.01)	(0.01)
Intercept	8.97^{***}	50594.28^{***}	0.35^{***}	0.26^{***}
	(0.09)	(270.61)	(0.00)	(0.00)
R^2	0.12	0.09	0.15	0.08
Num. obs.	12904	12904	12904	12904

Note: Table displays the linear regression estimates of officer characteristics on the average characteristics of the districts in which they work. Population and income are determined based on 2010 Census estimates. Violent crime rates are violent crimes in a month, based on Chicago City Data Portal crime data, per 10,000 population. The coefficients Recruit and Probationary are indicators for whether or not the officer is a new officer in their post-probationary period or a new officer in their probationary period, respectively. Robust standard errors are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Focusing on the main sample officers, Table A.6 displays regression estimates predicting the characteristics of the average unit a new officer serves in as a full officer. Notably, having more Black and Hispanic peers in one's academy cohort decreases the Black population share of the average district in which an officer works. As is evidenced from the table, Black officers are more likely to serve in Black districts, thus it is likely the case that having more Black officers in one's cohort reduces the probability of officers to be placed in Black districts.

However, this confounding assignment is unlikely to significantly bias estimates because new officers work in high crime areas with higher Black populations regardless of an officer's race, as shown in Table A.5. Despite this, main sample officers in the work in all of the 22 districts during the 2013-2018 period and no single district makes up more than 11% of assignments. Furthermore, it is evident from the very small amount of variation explained by observables in Table A.6 that there is likely much CPD-level demand choices made regardless of cohort observables and the actual influence of such observables are economically small. For example, a 5pp increase in cohort share Black only decreases a recruit's average district's Black population share by about 6% of the baseline mean. Given this, it is unlikely that the influence of cohort composition on new officers' assignments will significantly bias results, and controlling extensively for working environment, as in Section 3, should remove this bias completely.

Table A.6: Characteristics of Average District in Post Probationary Period

	Violent Crime	Median Income	Share Pop. White	Share Pop. Black	Share Pop. Hispanic
	(1)	(2)	(3)	(4)	(5)
Cohort Share Black	-11.12	10408.85	0.17	-0.92^{**}	0.70^{**}
	(8.44)	(9381.91)	(0.11)	(0.36)	(0.30)
Cohort Share Hispanic	-0.72	-3074.69	-0.01	-0.43^{*}	0.45^{**}
	(4.72)	(4902.40)	(0.07)	(0.23)	(0.18)
Cohort Share Other	7.20	54.20	-0.01	0.27	-0.22
	(8.98)	(7795.52)	(0.10)	(0.44)	(0.35)
Cohort Mean Age	-0.47	545.09	0.01	-0.00	-0.01
	(0.48)	(588.76)	(0.01)	(0.03)	(0.02)
Black	0.04	263.77	-0.01	0.09^{***}	-0.08^{***}
	(0.63)	(1019.01)	(0.01)	(0.03)	(0.02)
Hispanic	-0.18	76.50	0.00	-0.01	0.01
	(0.50)	(653.48)	(0.01)	(0.02)	(0.02)
Asian/Native American	0.09	550.50	0.05^{**}	-0.07	-0.01
	(1.60)	(1805.14)	(0.03)	(0.06)	(0.04)
Male	-0.11	373.01	0.01	0.00	-0.01
	(0.62)	(912.47)	(0.01)	(0.03)	(0.02)
Start Age	-0.13^{**}	106.82	0.00	-0.01^{**}	0.00^{***}
	(0.05)	(91.07)	(0.00)	(0.00)	(0.00)
Cohort Size	-0.04^{**}	22.48	-0.00	-0.00	0.00^{*}
	(0.02)	(18.07)	(0.00)	(0.00)	(0.00)
Intercept	37.60^{**}	12463.55	-0.11	1.20	-0.07
	(14.58)	(18605.44)	(0.19)	(0.73)	(0.60)
Mean Outcome	15.49	34203.262	0.085	0.7	0.184
R^2	0.02	0.01	0.01	0.04	0.05
Num. obs.	940	940	940	940	940

Note: Table displays the linear regression estimates of recruit and cohort characteristics on the average characteristics of the districts in which they work after their probationary period. Population and income are determined based on 2010 Census estimates. Violent crime rates are violent crimes in a month, based on Chicago City Data Portal crime data, per 10,000 population. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

A.5 Working Peers and Instructors

Another concern is that officers exposed to higher amounts of minorities in their cohorts may end up working with more minorities in the future. If contemporaneous peers influence arresting decisions, then the effect of academy peers may be capturing the selection of future minority peers and their influence. A similar concern is that academy diversity may influence the composition of one's field training officers. To test these, I regress the average composition of main sample officers' contemporaneous peers and field training officers on their cohort composition. Table A.7 displays these results.

In Columns (1) to (4), I use two main peer groupings, officers assigned to the same shift and watch number and the same sector (a subset of a district composed of multiple beats) or same beat, with Columns (1) and (3) being during the officer's probationary period and (2) and (4) being during the officer's time as a full officer. The dependent variable in Column (5) is the share of an officer's field training officers who are white during their probationary period. Relative to the mean white share across each group and officer type (probationary and full), the relationship between an officer's cohort share minority and their future peers' and FTOs' share white is not economically significant, and only one is statistically significant. The effect size with the largest magnitude (Column (3)) indicates that a 10pp increase in cohort share minority leads to a 0.016 decrease in the peer share white, which is a 3.4% decrease relative to the mean. The small and noisy, but consistently negative, relationship between cohort share minority and the share of white working peers is likely an artifact of this model not controlling for unit assignment: more minority peers slightly crowd out positions in high Black areas which also have slightly fewer white officers. But, as discussed in Appendix A.4, this effect is minor and almost all new officers go to high crime and high share Black districts. Given this, it is unlikely that the effect of future peer or training officer composition is driving the effects, and much of these small and noisy effects can be explained by the weak influence of cohort diversity on unit assignment. Furthermore, the extensive assignment controls discussed in Section 3 will absorb a significant amount of the minor differences in working peers as well.

While not studied in this paper, an officer's partner working with them in the same car or same beat on a specific day likely also influences their behavior. New officers are generally required to be placed with more experienced ones, so it is unlikely that cohort-mates work together. Furthermore, the working day rotation system creates a relatively chaotic environment during an officer's first few years in terms of partners. Essentially, who a new officer is able to work with is determined by the rotating day-off-group calendar, officers taking furlough days, and the fact that the units new officers are placed (high violent crime) are those that more senior officers tend to leave when vacancies are available elsewhere. So, for newer officers, who have the least choice in their shift, unit, and day-off-group, as well as no prior experience in the district (and thus no potential for having an established beat assignment), working-partners rotate frequently. Partners during these early years (focused on in this study) likely involve little endogenous selection relative to later on in their careers.

Table A.7: Average Characteristics of Peers and Training Officers

	Average Sector Share White		Average Beat Share White		FTO Share White
	Probationary	Full	Probationary	Full	Probationary
	(1)	(2)	(3)	(4)	(5)
Cohort Share Minority	-0.01	-0.13	-0.16^{**}	-0.09	0.06
	(0.07)	(0.11)	(0.08)	(0.17)	(0.15)
Cohort Mean Age	0.01^{*}	0.01^{**}	-0.00	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Black	-0.10^{***}	-0.15^{***}	-0.12^{***}	-0.29^{***}	-0.14^{***}
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
Hispanic	-0.03^{***}	-0.06^{***}	-0.05^{***}	-0.16^{***}	-0.04^{*}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Asian/Native American	0.04^{**}	-0.02	0.02	-0.11^{***}	0.05
	(0.02)	(0.02)	(0.04)	(0.03)	(0.05)
Male	-0.01	0.03^{***}	-0.00	0.05^{***}	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Start Age	0.00	-0.00^{**}	-0.00	-0.00^{**}	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cohort Size	-0.00	-0.00	-0.00	-0.00^{*}	-0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	0.22	0.22	0.68^{***}	1.05^{**}	0.94^{**}
	(0.18)	(0.24)	(0.25)	(0.44)	(0.46)
Mean Dep.	0.48	0.48	0.46	0.48	0.43
R^2	0.08	0.20	0.06	0.25	0.03
Num. obs.	940	940	940	940	940

Note: Table displays the linear regression estimates of main sample officer characteristics on their working peers and training officers during their probationary (odd columns) and post-probationary (even columns) periods. The dependent variable in Columns (1) and (2) is the share of an officer's peers who are white and working the same day, shift, and sector. The dependent variable in Columns (3) and (4) is the share of an officer's peers who are white and working the same day, shift, and beat number. The dependent variable in Column (5) is the share of an officer's field training officers who are white. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

A.6 Shrinkage Estimates

The individual fixed effects recovered by equation (2) are based upon finite observations of officers. This means that each estimated fixed effect will have some error associated with it, and it is crucial that measurement error is not driving the results. A common procedure to correct for this when using individual fixed effects, popularized by the teacher value added literature (Chetty, Friedman, and Rockoff (2014)), is to do an empirical Bayes shrinkage procedure (based on Morris (1983)). The idea is to shrink estimates (officer fixed effects) toward a prior mean based on how noisy the estimate is (high noise leads to a larger reduction in the estimate). Here, I will construct a shrunken estimate for each officer fixed effect ($\hat{\theta}_i$) based on how noisy it is ($Var[\hat{\theta}_i] = se(\theta_i)^2$) relative to the variance in the distribution of all fixed effects ($Var[\hat{\theta}] = \frac{1}{N}\sum_i^N \hat{\theta}_i^2$):

$$\hat{\theta}_i^{shrunken} = \hat{\theta}_i * \frac{Var[\hat{\theta}]}{Var[\theta_i] + Var[\hat{\theta}]}.$$

This is derived from the posterior mean of normal distribution with prior mean equal to zero being:

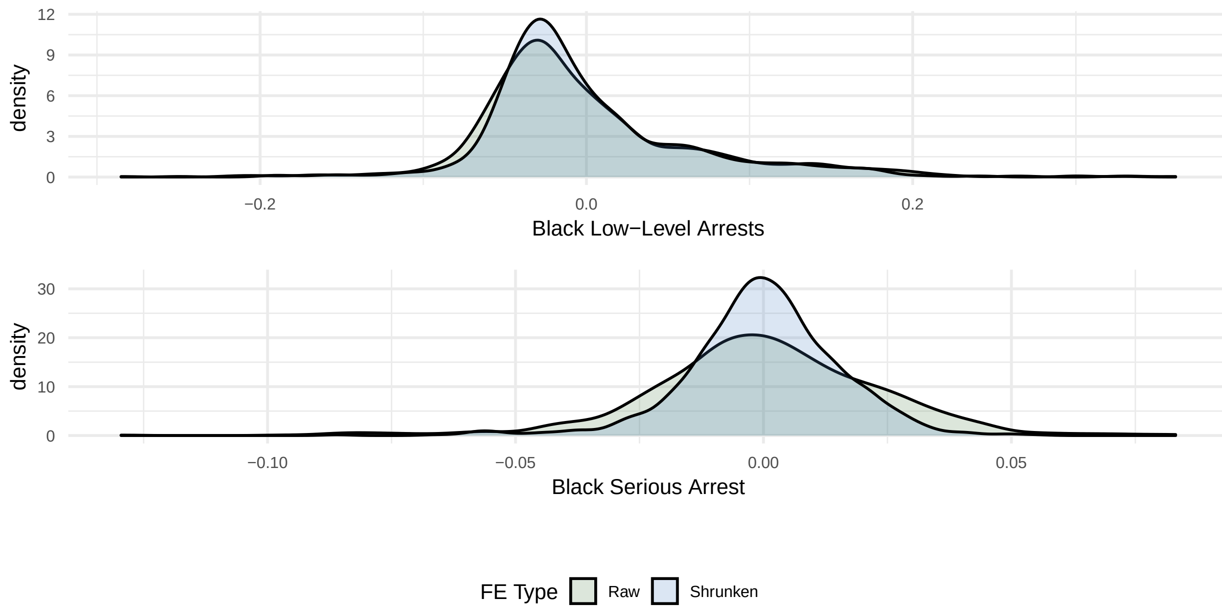
$$\theta_i^n = \bar{\theta}_i \frac{\sigma^2}{\sigma^2 + \frac{\sigma_i^2}{n}},$$

where θ_i is drawn from a $N(0, \sigma^2)$ and each observation of $\theta_i^t = \theta_i + \epsilon_i^t$, where $\epsilon_i^t \sim N(0, \sigma_i^2)$. $\bar{\theta}_i$ can be seen as the fixed effect estimate, $\frac{\sigma_i^2}{n}$ is the $se(\bar{\theta}_i)^2$, and σ^2 as the variance across estimates of θ_i's.[50]

As expected, due to the relatively large number of observations per officer in my sample (the median main sample officer has 575 observations and the median full sample officer has 401 observations), the shrunken fixed effects are similar to those of the main results. Figure A.2 displays the distribution of raw and shrunken fixed effects for arresting Blacks for low-level and serious crimes. For low-level arrests, which are more common, the raw and shrunken distributions are more similar.

[50] A key advantage of the 'lfe' R package is its ability to estimate standard errors for fixed effects via bootstrapping. However, as discussed in Hahn and Liao (2021), bootstrapped standard errors tend to be conservative; if this is the case in my environment, I may be 'over'-shrinking because overly conservative (large) standard errors will lead to smaller rescaling factors. However, this is unlikely to be a concern as the results are consistent with raw fixed effects as well as post-shrinking fixed effects.

Figure A.2: Distributions of Raw and Shrunken Main Sample Fixed Effects



Note: The figure compares the centered distributions of main sample officer fixed effects for arrests of Blacks for low−level and serious crimes, recovered from equation (2). Raw fixed effects are unaltered, while shrunken fixed effects were subject to the Bayesian shrinkage procedure described in Appendix A.6.

A.7 Randomization-Based Inference

Randomization inference (or randomization-based inference) allows us to construct an empirical distribution of coefficients under the null hypothesis, that peers have no effect on officer fixed effects. This is preferable to traditional asymptotic inference in which the error in estimates is a result of sampling error because in such environments, there is no sampling error; e.g., the sample of CPD recruits between 2009 and 2016 is the population. Such methods have their origin in Fisher (1925), wherein one wants to test to see if they can reject the 'sharp' null hypothesis that the treatment has no effect on the outcome of interest, and much of this section will follow Athey, Eckles, and Imbens (2018). Let us generalize equation (3) (removing superscript k for simplicity) as a potential outcomes function:

$$\theta_i(P_i = \overline{X}_{c(i)}) = \alpha_{p(i)} + \pi_1 \overline{X}_{c(i)} + \pi_2 X_i + v_i$$

Then the potential outcomes function for an individual, θ_i, takes in a value for i's peer composition P and tells us what the individual's fixed effect for arrests would be had they had peer composition P in the academy. As discussed in Athey, Eckles, and Imbens (2018), under a sharp null hypothesis of no effect, given some treatment assignment P' and the realized outcomes for that specific assignment $\theta_i(P')$, one can infer the value of the outcome at any other treatment assignment. Essentially if under the null that $\pi_1 = 0$, then $\theta_i(P) = \theta_i(P') \forall P, P' \in \mathbb{P}$ where P' is any possible peer composition and $\mathbb{P}$ is the space of all possible treatment (peer) assignments. The intuition is that if the true peer effect is zero ($\pi_1 = 0$), then it should not matter what treatment (peer composition) is assigned.

Now, we can test this null hypothesis. We can generate test statistics based on the distribution of estimated treatment effects (π_1^r), the 'randomization distribution', when the treatment status is randomly assigned. With this distribution of estimated peer effects under the randomized treatments, we compare the estimate from our actual data ($\hat{\pi}_1$) to the randomization distribution and recover the p-value– the likelihood of finding an effect more extreme than the one estimated under the null hypothesis that treatment has no effect. Again, borrowing from Athey, Eckles, and Imbens (2018):

$$p\text{-}value = Pr(|\hat{\pi}_1(\theta_i(P = \overline{X}_{c(i)}))| \geq |\pi_1^r(\theta_i(P'))|)$$

With this p-value, we can assess the likelihood that the estimate recovered from the actual data ($\hat{\pi}_1$) is consistent with the null hypothesis, that the the peer effect is null.

In practice, constructing the randomization distribution can be done in two ways. (1) (Re-assigning Treatment) Randomly re-assigning individuals to cohorts within exams and ensuring cohort sizes remain the same and thus constructing randomized treatments ($\bar{X}^r_{c^r(i)}$), then estimate:

$$\hat{\theta}_i = \alpha_{p(i)} + \pi_1^r \overline{X}^r_{c^r(i)} + \pi_2 X_i + v_i$$

Or (2) (Re-assigning Outcomes), randomly re-assigning outcomes to individuals (θ_i^r):

$$\hat{\theta}_i^r = \alpha_{p(i)} + \pi_1^r \overline{X}_{c(i)} + \pi_2^r X_i + v_i$$

In either case, this procedure can be repeated N number of times (I perform 1,000 iterations for method (1) and 5,000 iterations for method (2)) with each iteration producing an estimate of π_1^r. Then, the coefficient using the actual data, $\hat{\pi}_1$ can be compared with the distribution of $\hat{\pi}_1^r$ to obtain a p-values as discussed above. Method (1) is used in Caeyers and Fafchamps (2016) and Michelman, Price, and Zimmerman (2021), while Method (2) is used in Carrell, Sacerdote, and West (2013) and Carrell, Hoekstra, and West (2019).[51]

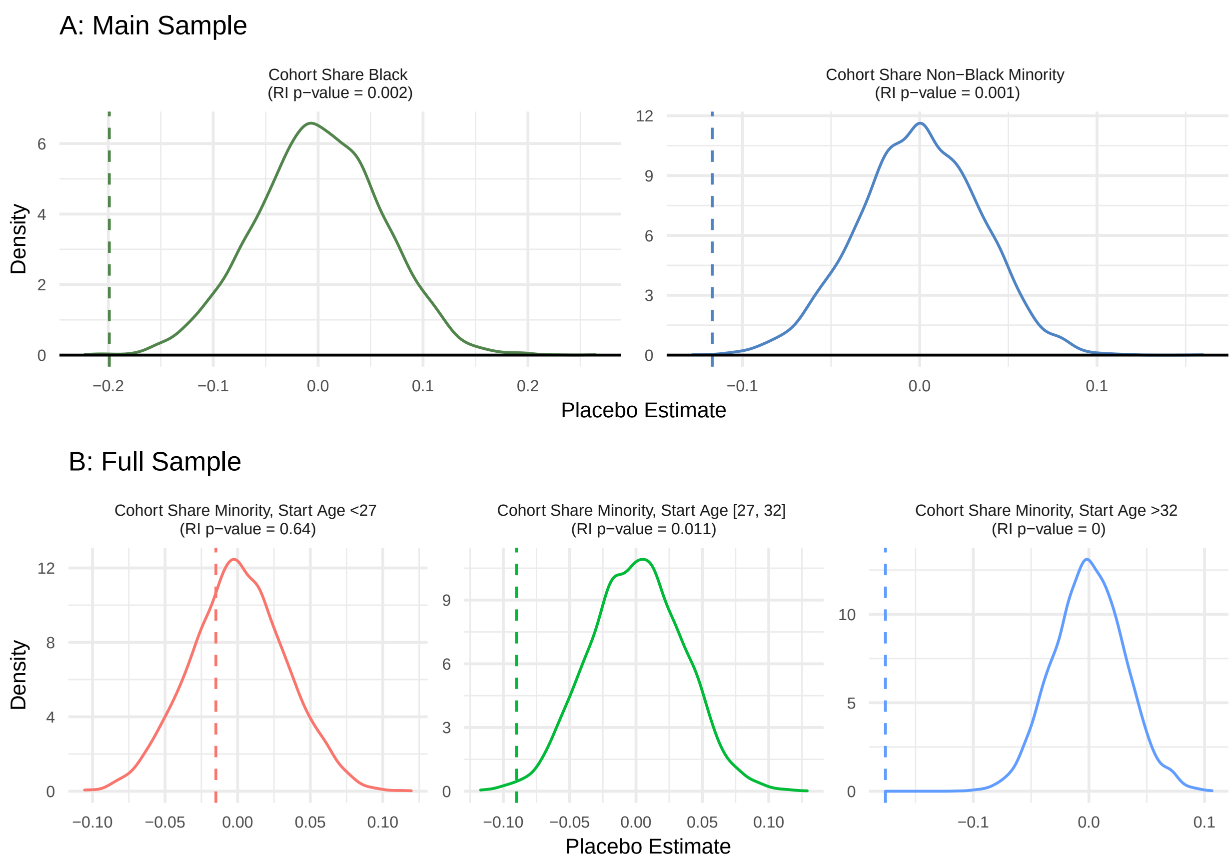
Effectively, method (1) differs from (2) in that (1) randomizes only the treatment of interest while (2) also scrambles the relationship between the controls (X_i) and the outcome.

[51] The reassignment of outcomes to observations is based on the author's understanding of the replication code for Carrell, Hoekstra, and West (2019), see lines 529 to 560 for reference in 20170069_main.do in their replication files.

In practice because of sample attrition, method (1) involves re-drawing cohorts (within exams) using recruits in the final sample and those who are dropped from it, where as method (2) randomizes the outcomes (within exams) between officers who appear in the final sample.[52] In both cases, two-sided p-values are computed by ranking the coefficient in the main results within the distribution of placebo coefficients.

As shown in Figures 2 and A.3, the estimated peer effects which are reported as statistically significant in the main results are statistically significant using either method, with empirical p-values generally smaller than those in the main results.

Figure A.3: Randomization Distribution of Coefficients (Reassigned Fixed Effects)



Note: Figure visualizes the distribution of coefficients estimated using 5,000 placebo rounds to conduct randomization inference, as discussed in Appendix A.7. Each placebo round involves randomly re−assigning the outcome variable to an officer within a exam period. Coefficients are the effects of cohort shares Black and non−Black minorities in the main sample (panel A) and the effects of cohort share minority by age group for full sample officers (panel B) on the outcome variable, shrunken officer fixed effects for arresting Blacks for low−level crimes. The dashed vertical lines correspond the coefficient estimated in the main specification (actual cohorts), and the RI p−value denotes the p−value resulting from a two−tailed test which ranks the magnitude of the actual coefficient among the magnitudes of the 5,000 placebo coefficients. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Coefficients estimated using equation (3) using full controls (specification in Column (13) in Table 3 for main sample and specification in Column (5) in Table (4) for the full sample).

[52] Both methods do take the error in the outcome (i.e., $\hat{\theta}_i$ is an estimate with measurement error) as given.

A.8 Small Class Effects (Homerooms)

While many classes were composed of almost all the officers in one's cohort, smaller sub-cohort groups ("homerooms") are identifiable when restricting to classes with fewer than 30 recruits. I use data on individual classes the officers took while in the academy to see if recruits in small group (homeroom) composition is driving the effects of cohort composition on the outcomes. If this is the case, then it is more likely that instructor effects are a contributing factor.

The training data provided lists the set of classes each probationary officer took during their time at the academy. This includes classes on the data base access, report writing, terrorism, chemical and radioactive events, and use of force. Many classes are large containing almost all (or a large portion) of a cohort's members. A subset of courses contain fewer officers per class, meaning there is larger within-cohort variation on which cohort members attended these courses together.

I use the set of trainings during the academy that full sample officers took which had fewer than 30 officers attend and a sufficiently high share of the classes being from the same cohort. With this set of courses, I created a weighted undirected network of recruits within cohorts and use the "edge betweenness" clustering algorithm (Newman and Girvan (2004)) in order to partition these networks into sub-communities of officers that had the strongest ties based on classes taken together. I refer to these sub-cohorts as homerooms.[53]

After some filters, the final sample of officers in the homerooms (also in the full sample) is 2,093 in 105 homerooms. Not all recruits are present in the final homeroom data due to matching issues and filters (89.6% of full sample officers are in the final homeroom data) and I restrict to homerooms with between 14 and 30 recruits. Due to the smaller size of these homerooms, there is much more variation in compositions. For example, there is 2.5 times more variation in cohort share minority for homerooms relative to cohorts. Nevertheless, homeroom and cohort compositions are highly correlated: For example, the correlation between cohort share Black and homeroom share Black is 0.77.

To see if homeroom composition is driving the results, I re-estimate equation (3) on the full sample using homeroom instead of cohort composition. The results are displayed in Table A.8. The effects of homeroom composition are similar to those for cohort composition (Columns (1) and (2)), while adding cohort fixed effects to see if variation between homerooms within cohorts has an effect produces statistically and economically insignificant results (Columns (3) and (4)). As noted in Section 6.2, this is not consistent with instructor

[53]The edge betweenness clustering algorithm is implemented in the igraph package in R (Csardi and Nepusz (2005)).

effects driving the results.

Table A.8: Effect of Homeroom Diversity on Arrest Propensity

	Shrunken Arrest Propensity			
	Black Low-Level	Black Serious	Black Low-Level	Black Serious
	(1)	(2)	(3)	(4)
Homeroom Share Minority, Start Age <27	−0.021	0.006	−0.014	0.001
	(0.029)	(0.004)	(0.015)	(0.004)
Homeroom Share Minority, Start Age [27, 32]	−0.031	0.005	0.021	0.007^{*}
	(0.026)	(0.004)	(0.015)	(0.004)
Homeroom Share Minority, Start Age >32	-0.073^{***}	0.012^{***}	−0.007	0.004
	(0.025)	(0.004)	(0.017)	(0.004)
Homeroom Share White, Start Age [27, 32]	-0.038^{*}	0.001	−0.003	−0.002
	(0.022)	(0.004)	(0.014)	(0.004)
Homeroom Share White, Start Age >32	−0.042	0.007	0.018	0.007
	(0.031)	(0.005)	(0.021)	(0.006)
Controls	Full	Full	Full	Full
Cohort FE			X	X
R^2	0.433	0.141	0.488	0.165
Num. obs.	2093	2093	2093	2093

Note: Table displays the effect of homeroom composition on full sample officers' propensities to make arrest of Blacks for low-level and serious crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Homerooms are sub-cohorts constructed using individual class training data as described in Appendix A.8. Homeroom shares are computed as the leave-out mean of the officer's homeroom's initial composition. Full controls refers to the additional controls in Column (13) of Table 3 with homeroom size substituted for cohort size. Cohort FEs refer to cohort-specific fixed effects. Standard errors clustered at homeroom level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

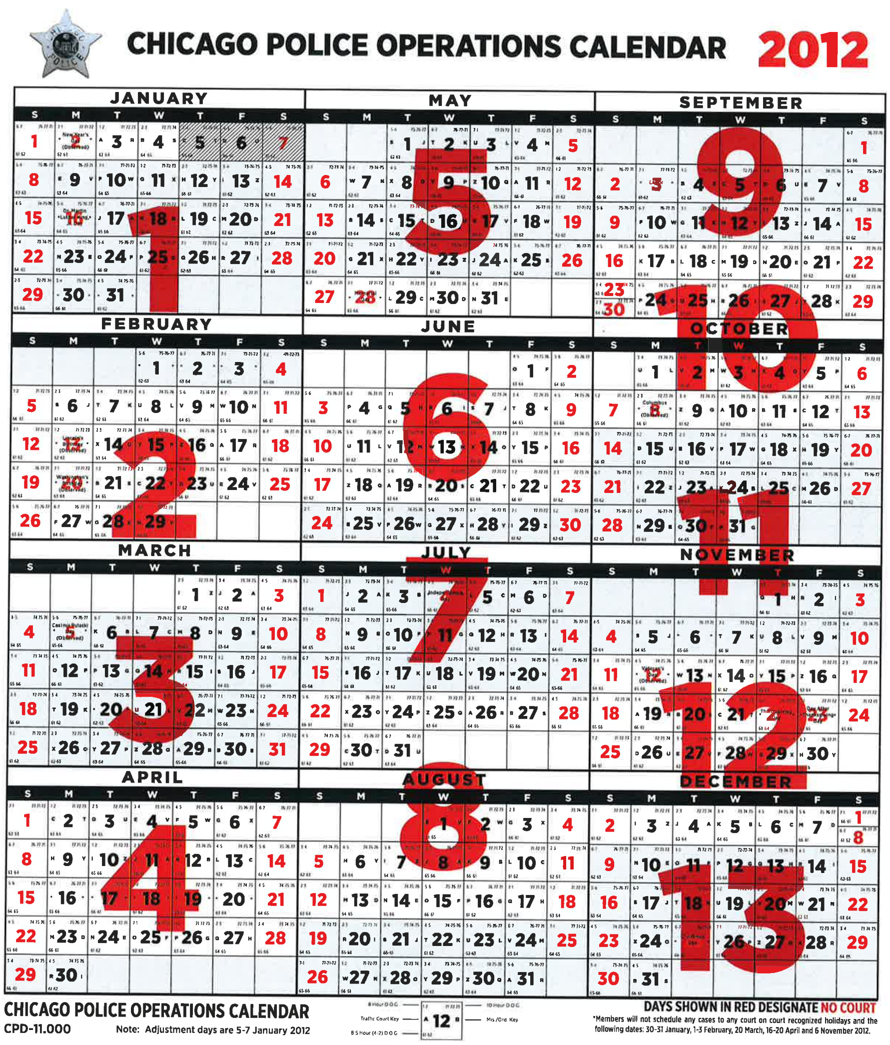
B Appendix B - Additional Figures and Tables

Figure B.1: CPD Exam Information

Exam	Date of administration	Attended	Passed	Failed
Police Entrance 1999	3/15/1999; 3/16/1999	3,967	No info available	No info available
Police Entrance 1999	1/5/2000	2,517	No info available	No info available
Police Entrance 2000	7/1/2000	2,053	No info available	No info available
Police Entrance 2000	1/4/2001	1,829	No info available	No info available
Police Entrance 2001	5/19/2001	1,923	No info available	No info available
Police Entrance 2002	1/12/2002	3,150	No info available	No info available
Police Entrance 2003	11/22/2003	3,875	No info available	No info available
Police Entrance 2004	11/20/2004	4,163	No info available	No info available
Police Entrance 2005	2/18/2006; 2/19/2006	4,061	3,338	723
Police Entrance 2006-1	6/4/2006	1,508	1,255	253
Police Entrance 2006-2	8/6/2006	1,025	863	162
Police Entrance 2006-3	11/5/2006	1,795	1,487	308
Police Entrance 2010	12/11/2010	8,621	7,689	932
Police Entrance 2010 make up	makeups: 3/12/2011; 6/11/2011; 9/25/2011; 12/3/2011; 6/2/2013; 12/1/2012; 3/9/2013	No info available	No info available	No info available
Police Entrance 2013	12/14/2013 & military makeups (6/28/2014; 12/7/2014; 6/13/2015; 12/6/2015)	14,788	12,877	1,911
Police Entrance 2016	4/16/2016 & make ups :12/3/2016; 12/4/2016	10,199	9,023	1,176
Police Entrance Spring 2017	4/1/2017-4/2/2017	8,620	7,437	1,183
Police Entrance Winter 2017	12/16/2017,12/17/2017 & makeup: 2/24/2018	7,294	6,418	876
Police Entrance Spring 2018	5/5/2018 & 5/6/2018 & makeup: 6/23/2018	4,273	3,789	484
Police Entrance Winter 2018	12/8/2018	4,433	3,964	469
Police Entrance Winter 2018 make up	3/9/2019	Hasn't occurred	N/A	N/A

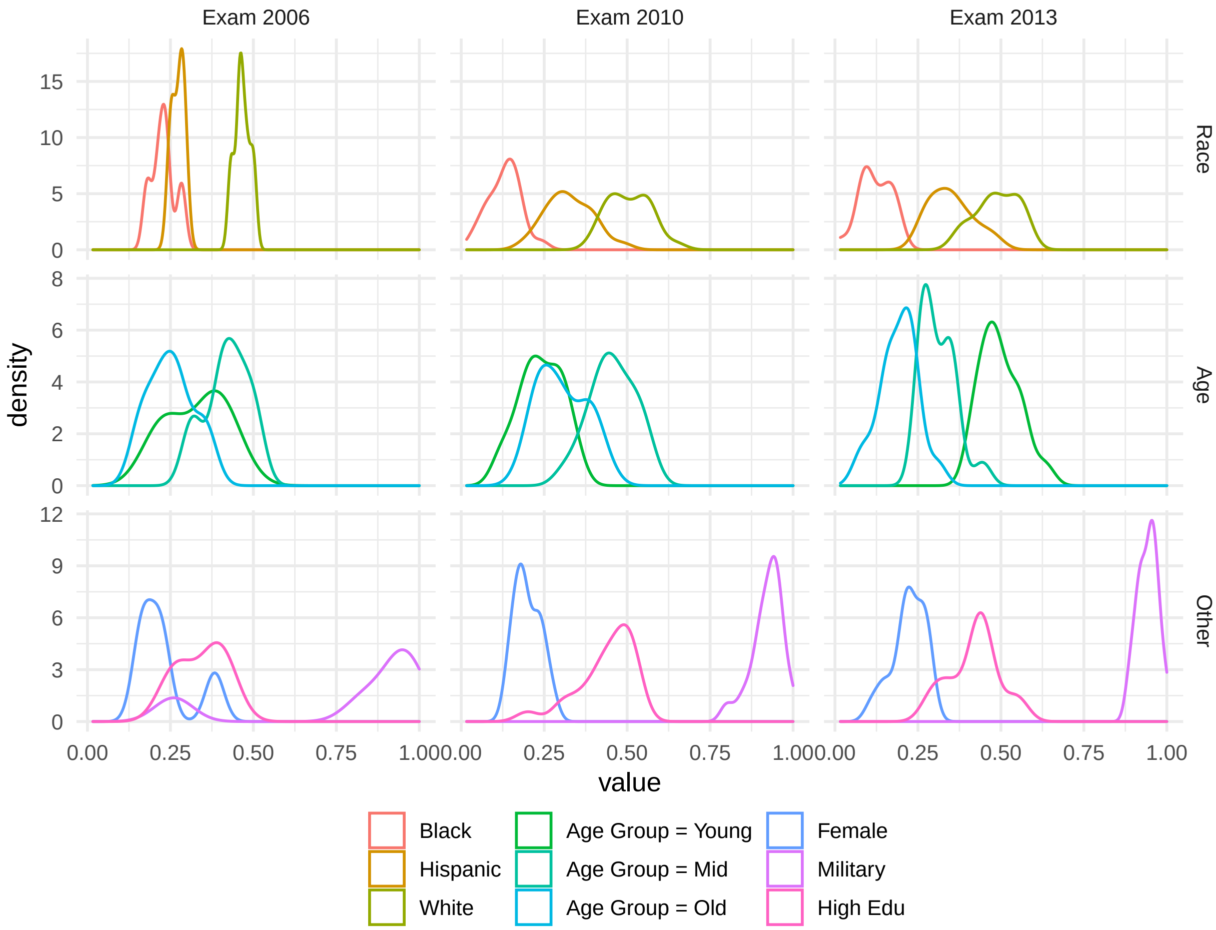
Note: Figure displays information on CPD entrance exam information, the date of the exam and the numbers of applicants that attended, passed, and failed the exam.

Figure B.2: CPD Operations Calendar (2012)



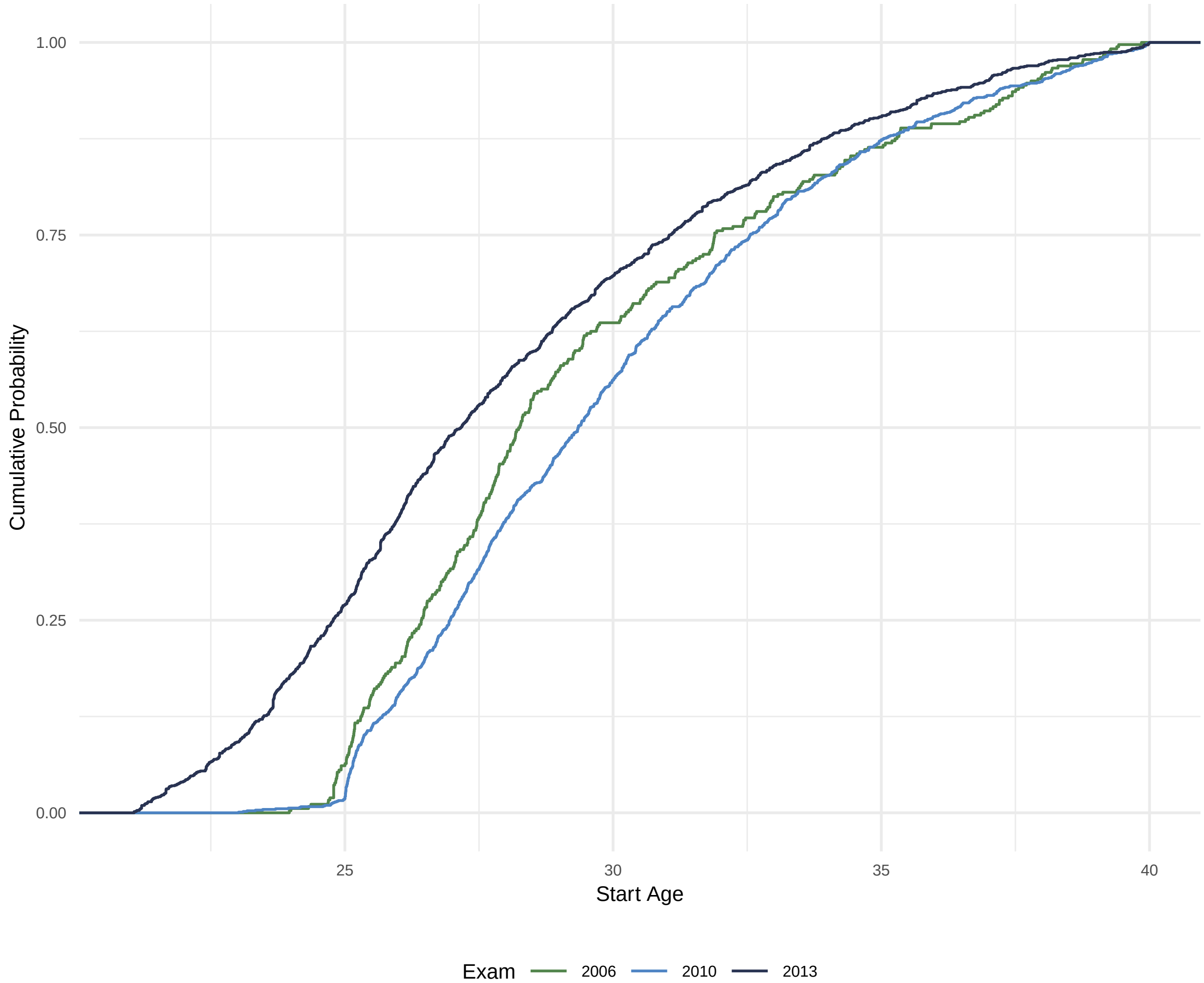
Note: Figure displays an example of the CPD operations calendar for the year 2012.

Figure B.3: Cohort Composition



Note: Figure displays the distributions of cohort compositions for Exam periods 2010 (main sample), 2006, and 2013 for characteristics including race (share Black, Hispanic, white), age (young = <27, mid=[27,32], and old= >32), gender (share female), and shares of those with military experience and high education (Bachelors or above).

Figure B.4: CDF of New Officer Start Ages



Note: Figure displays the cumulative distributions of officer start ages in cohorts for each Exam 2006, 2010 (main sample), and 2013. The figure illustrates that officers cannot begin at the academy after the age of 40 or before the age of 23 prior to Exam 2013 and 21 in Exam 2013 due to a policy change.

Table B.1: Association between Officer Characteristics and Arrest Propensity (per 100 shifts)

	Minority Arrests	Non-Black Arrests	Black Low-Level Arrests	Black Serious Arrests
	(1)	(2)	(3)	(4)
Black	-3.56^{***}	-1.08^{***}	-2.83^{***}	0.08
	(0.64)	(0.24)	(0.50)	(0.23)
Hispanic	-1.71^{***}	-0.15	-1.54^{***}	-0.13
	(0.55)	(0.22)	(0.38)	(0.19)
Asian/Native American	-2.48^{***}	-0.27	-1.91^{***}	-0.42^{*}
	(0.85)	(0.39)	(0.63)	(0.24)
Minority, Start Age [27, 32]	-1.81^{***}	-0.17	-1.20^{**}	-0.54^{***}
	(0.70)	(0.21)	(0.51)	(0.16)
Minority, Start Age >32	-1.87^{***}	-0.34^{*}	-1.36^{***}	-0.28
	(0.49)	(0.20)	(0.39)	(0.17)
White, Start Age [27, 32]	-1.53^{***}	-0.24	-1.21^{***}	-0.11
	(0.55)	(0.21)	(0.42)	(0.12)
White, Start Age >32	-3.75^{***}	-0.47	-2.91^{***}	-0.45^{**}
	(0.69)	(0.30)	(0.49)	(0.18)
Minority x Female	-2.98^{***}	-0.49^{***}	-1.90^{***}	-0.58^{***}
	(0.48)	(0.15)	(0.35)	(0.16)
White x Female	-3.08^{***}	-0.43	-2.17^{***}	-0.66^{***}
	(0.57)	(0.29)	(0.44)	(0.18)
Military	-1.60	-0.48	-1.54	0.36
	(1.08)	(0.39)	(1.02)	(0.41)
High Edu	0.44	0.07	0.44	-0.02
	(0.43)	(0.14)	(0.34)	(0.12)
Exam 2010	-7.34^{***}	-1.88^{***}	-9.19^{***}	3.33^{***}
	(0.93)	(0.22)	(1.03)	(0.32)
Exam 2013	-15.32^{***}	-3.76^{***}	-18.62^{***}	6.22^{***}
	(1.01)	(0.24)	(1.16)	(0.35)
Intercept	-101.24^{***}	-79.62^{***}	-65.59^{***}	24.91^{***}
	(1.31)	(0.47)	(1.31)	(0.52)
R^2	0.32	0.16	0.51	0.47
Num. obs.	2336	2336	2336	2336

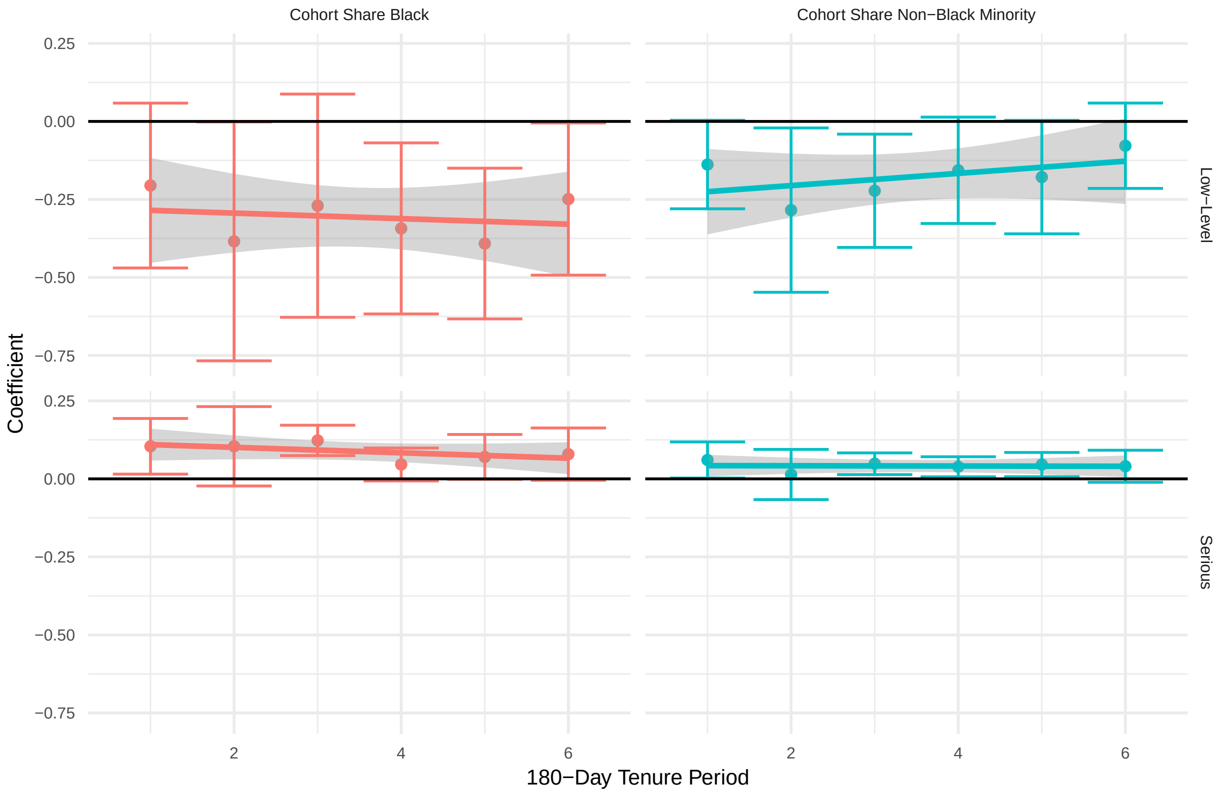
Note: Table displays the linear regression estimates of full sample officer characteristics on their individual propensities to make arrests of various types, recovered using equation 2. High edu corresponds to having a Bachelors degree or above. Non-Black includes white, Hispanic, Asian, Native American, etc.. Coefficients are scaled up to be per 100 shifts for easy of interpretation. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2: Effect of Cohort Diversity on Low-Level Arrest Propensity Across Arrestee Groups - Main Sample

	All	Minority	Black	Non-Black	Non-Black Minority	White
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	-0.34^{**}	-0.31^{**}	-0.26^{**}	-0.09^{*}	-0.06	-0.03^{*}
	(0.16)	(0.15)	(0.12)	(0.05)	(0.04)	(0.02)
Cohort Share Non-Black Minority	-0.21^{*}	-0.19^{**}	-0.15^{**}	-0.06	-0.04	-0.02
	(0.11)	(0.10)	(0.07)	(0.04)	(0.03)	(0.01)
Black	-0.03^{***}	-0.03^{***}	-0.02^{***}	-0.01^{***}	-0.01^{***}	-0.00^{**}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Hispanic	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.00	-0.00	-0.00^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Asian/Native American	-0.02	-0.02	-0.02^{**}	-0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Male	0.02^{***}	0.02^{***}	0.02^{***}	0.00	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Start Age	-0.00^{***}	-0.00^{***}	-0.00^{***}	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cohort Size	0.00	0.00	0.00	0.00	0.00	0.00^{*}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	-1.59^{***}	-1.34^{***}	-0.66^{***}	-0.93^{***}	-0.68^{***}	-0.25^{***}
	(0.07)	(0.06)	(0.05)	(0.02)	(0.01)	(0.01)
R^2	0.08	0.08	0.08	0.03	0.02	0.04
Num. obs.	940	940	940	940	940	940

Note: Table displays the effect of cohort composition on main sample officer outcomes. The outcomes are individual officer fixed effects, estimated using equation (2), for all arrests, minority, Black, non-Black, non-Black minority (mainly Hispanic), and white; effects estimated using equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Figure B.5: Dynamic Effect of Peer Composition on Propensity to Arrest Blacks



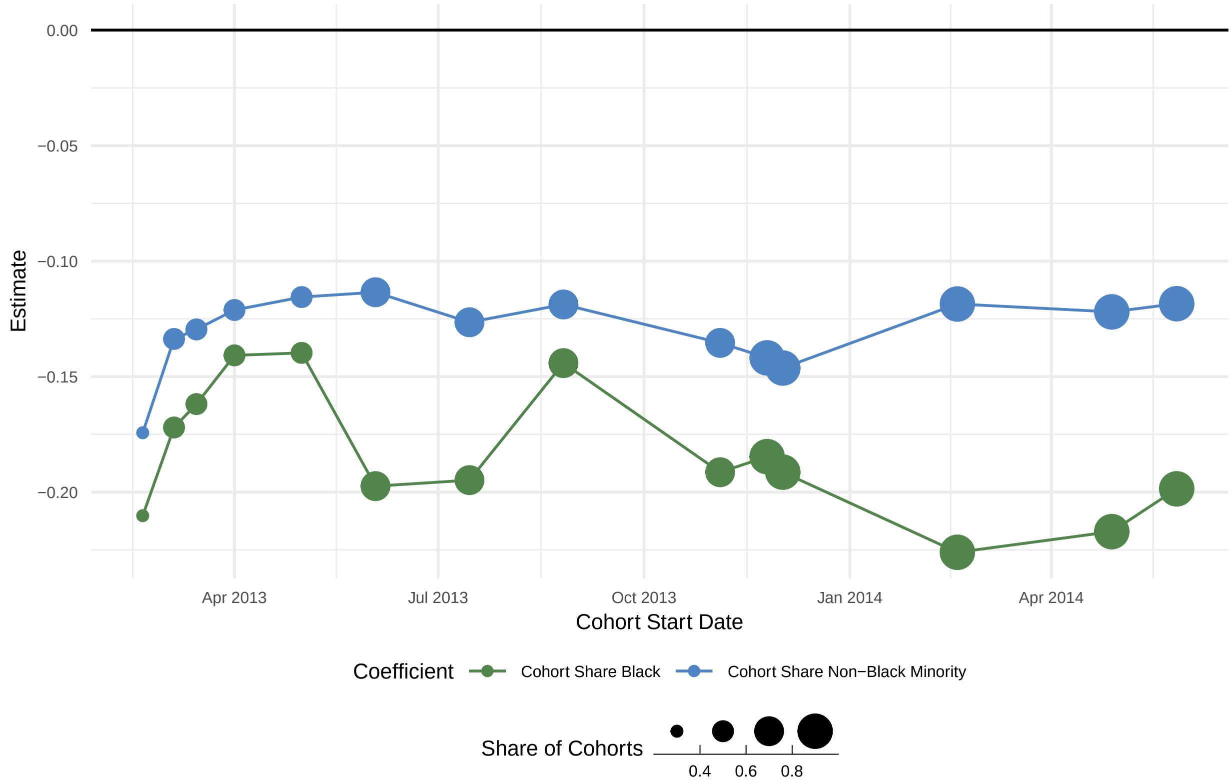
Note: Figure visualized the effects of cohort shares Black, Hispanic, and other non−whites on main sample officer fixed effects for arresting Blacks for serious and low−level crimes over their careers. Unshrunken (raw) fixed effects were recovered by modifying equation 2 such that officers had a separate fixed effect computed for each 180 day period of their tenure. Coefficients are estimated using equation 3 and the main specification (see Columns (9) and (10) in Table 3) separately for each 180−day period fixed effect. Cohort shares are computed as the leave−out mean of the officer's cohort's initial composition. Grey error bars indicate 95 percent confidence intervals based on standard errors clustered at cohort level.

Table B.3: Decomposed Effect of Cohort Composition on Low-Quality Arrest Propensity

	Shrunken Low-Level Black Arrest Propensity					
	Main Sample			Full Sample		
	Not Guilty	Dropped	Missing	Not Guilty	Dropped	Missing
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	-0.049	-0.044^{***}	-0.011			
	(0.031)	(0.014)	(0.013)			
Cohort Share Non-Black Minority	-0.033^{*}	-0.029^{**}	-0.006			
	(0.017)	(0.012)	(0.006)			
Cohort Share Minority, Start Age <27				-0.001	-0.015	0.009
				(0.014)	(0.010)	(0.007)
Cohort Share Minority, Start Age [27, 32]				-0.021	-0.029^{**}	-0.002
				(0.019)	(0.012)	(0.010)
Cohort Share Minority, Start Age >32				-0.047^{***}	-0.048^{***}	0.009
				(0.012)	(0.009)	(0.008)
Controls	Full	Full	Full	Full	Full	Full
R^2	0.061	0.057	0.032	0.443	0.497	0.049
Num. obs.	940	940	940	2336	2336	2336

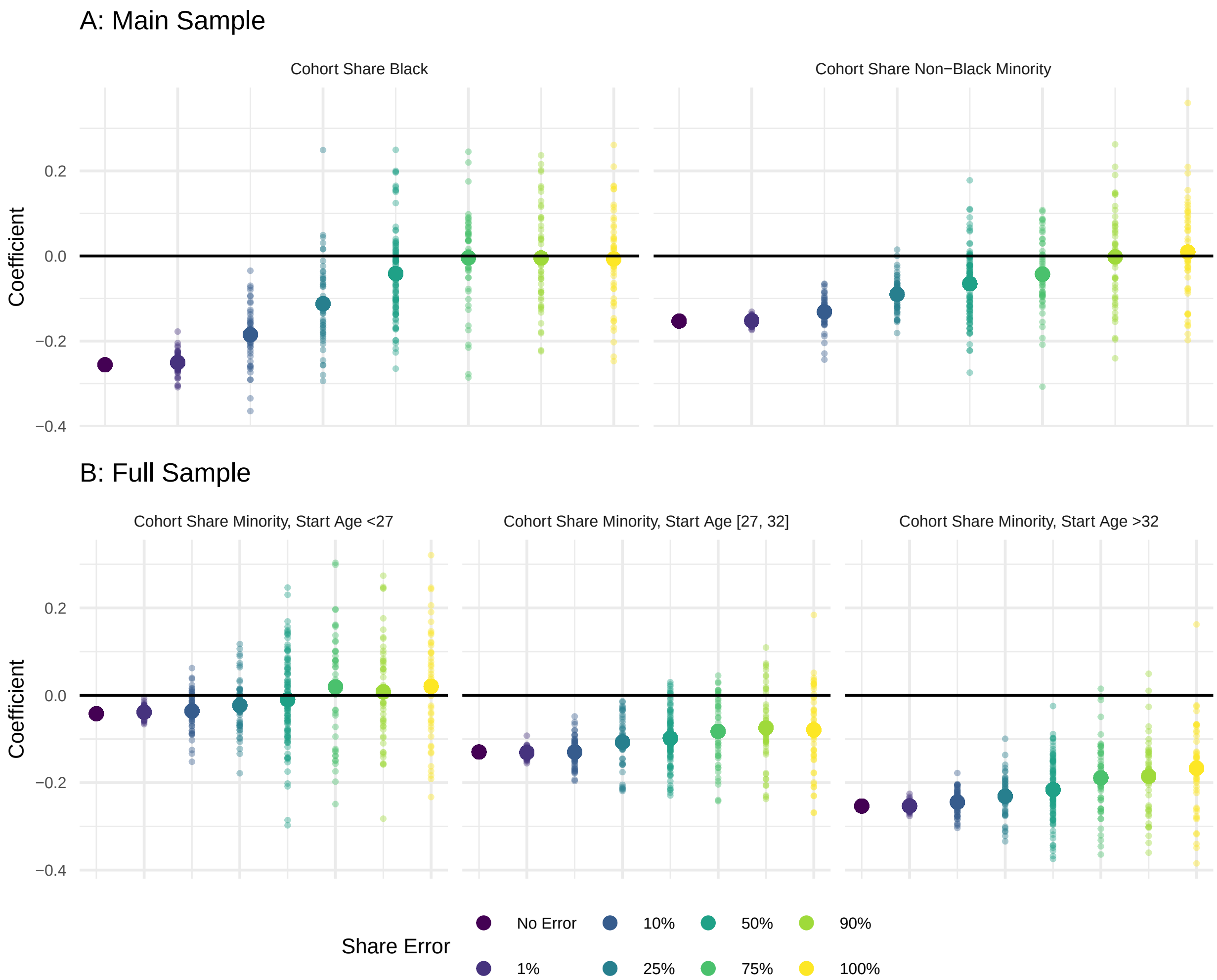
Note: Table displays the effect of cohort composition on main and full sample officers' propensities to low (non-guilty) quality arrest of Blacks for low-level crimes decomposed into not guilty (finding of not guilty), dropped (case dropped/dismissed), and missing (case does not appear in court data). Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Figure B.6: Coefficient Estimates by Increasing Number of Main Sample Cohorts



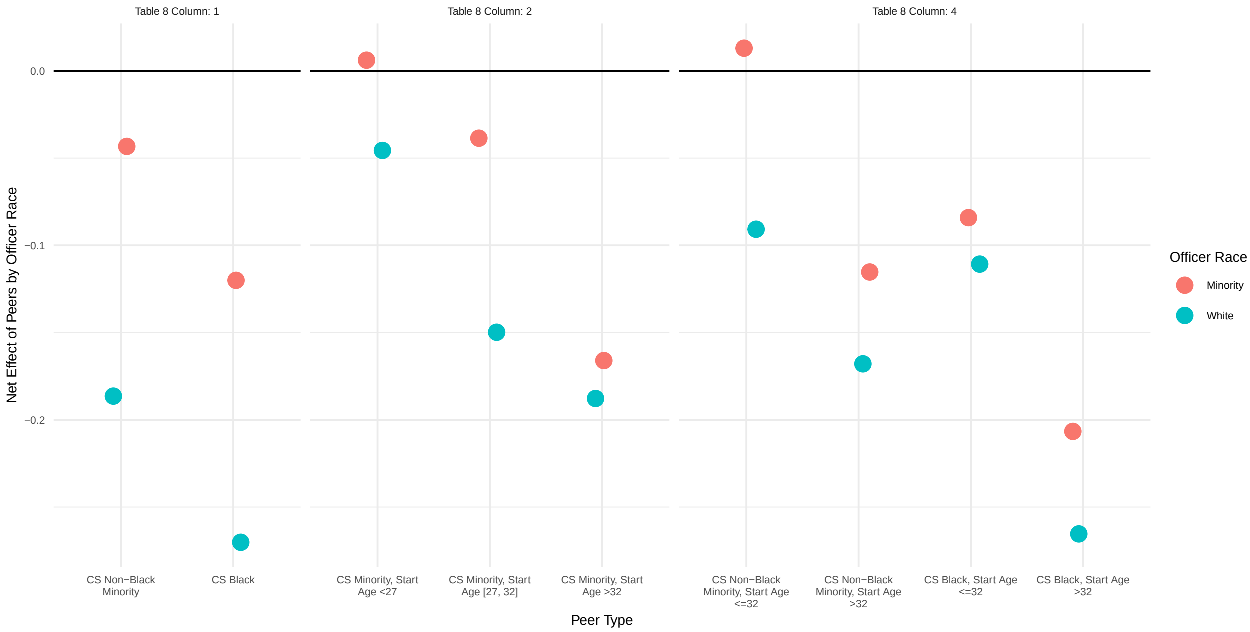
Note: Figure displays coefficients (y−axis) recovered from estimating equation 3 on the main sample of cohorts which started on or before each date (x−axis). The dependent variable is shrunken officer propensity to arrest Blacks for low−level crimes. As the sample size increased (more cohorts are included), the coefficients become more stable.

Figure B.7: Change in Coefficients with Measurement Error



Note: Figure visualizes how coefficients change as measurement error is added to peer racial composition. Coefficients are the effects of cohort shares Black and non−Black
sample (panel A) and the effects of cohort share minority by age group for full sample officers (panel B) on shrunken officer fixed effects for arresting Blacks for low−level crim
error is induced by taking the initial sample and assigning racial group (Black, Non−Black minority, and white) based on a uniform random variable for some share ('Share Err
peer compositions are computed. For each share error of observations with measurement error, this exercise is repeated 50 times, and each faint dot corresponds to a particu
(on per value of share error), are the mean coefficients across runs. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as descr
Coefficients estimated using equation (3) using full controls (specification in Column (13) in Table 3 for main sample and specification in Column (5) in Table (4) for the full sar

Figure B.8: Visualization of Interaction Terms (Table 8)



Note: Figure visualized the net effects of cohort shares Columns 1,2, and 4 in Table 8 by sample (main and full) for white officers and minority officers. White officer effects are computed by adding the minority officer coefficient to the white officer interaction coefficient

Table B.4: Effect of Cohort Diversity on Arrest Propensity Excluding Peer Officers

	Shrunken Low-Level Black Arrest Propensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	-0.271^{**}	-0.168^{**}				
	(0.107)	(0.076)				
Cohort Share Non-Black Minority	-0.187^{**}	-0.093				
	(0.079)	(0.061)				
Cohort Share Minority, Start Age >32			-0.134^{***}	-0.127^{**}		
			(0.052)	(0.062)		
Cohort Share White, Start Age >32			-0.057	-0.132^{*}		
			(0.084)	(0.080)		
White x Cohort Share Minority, Start Age >32				-0.008		
				(0.059)		
White x Cohort Share White, Start Age >32				0.134^{**}		
				(0.060)		
Cohort Share Minority and Female					-0.200^{**}	-0.165^{**}
					(0.080)	(0.071)
Cohort Share White and Female					-0.034	0.011
					(0.101)	(0.104)
White x Cohort Share Minority, Female						-0.067
						(0.065)
White x Cohort Share White, Female						-0.081
						(0.093)
Excluded Officers	Minority	Minority	Start Age > 32	Start Age > 32	Female	Female
Sample	Main	Full	Full	Full	Full	Full
Controls	Full	Full	Full	Full	Full	Full
R^2	0.068	0.376	0.400	0.402	0.385	0.385
Num. obs.	482	1161	1756	1756	1841	1841

Note: Table displays the effect of cohort composition on main (Column (1)) and full sample (Columns (2)-(4)) officers' propensities to make arrest of Blacks for low-level crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Each column excludes the group of officers which comprise the peer group whose effect is estimated. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3 for main sample with the addition of exam fixed effects for full sample regressions, and it excludes any control which is colinear with the excluded group (e.g., controls for gender are excluded from Column (4)). Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.5: Alternative Mechanisms for Effect of Peer Diversity on Arrest Propensity

	Shrunken Arrest Propensity Black Low-Level			
	Full Sample			
	(1)	(2)	(3)	(4)
Cohort Share Minority, Start Age <27	0.028	0.025	-0.362^{***}	-0.365^{***}
	(0.072)	(0.072)	(0.130)	(0.132)
Cohort Share Minority, Start Age [27, 32]	-0.041	-0.043	-0.421^{**}	-0.424^{**}
	(0.086)	(0.086)	(0.178)	(0.180)
Cohort Share Minority, Start Age >32	-0.150^{***}	-0.151^{***}	-0.510^{***}	-0.510^{***}
	(0.049)	(0.049)	(0.146)	(0.147)
Cohort Share White, Start Age >=27	-0.066^{*}	-0.067^{*}	-0.066^{*}	-0.068^{*}
	(0.039)	(0.039)	(0.038)	(0.038)
Cohort Share Minority and High Edu	-0.085	-0.093		
	(0.058)	(0.068)		
Cohort Share White and High Edu	0.019	-0.031		
	(0.053)	(0.051)		
White x Cohort Share Minority and High Edu		0.011		
		(0.044)		
White x Cohort Share White, High Edu		0.101^{***}		
		(0.037)		
Cohort Share Minority and Military			0.161^{*}	0.179^{*}
			(0.088)	(0.093)
Cohort Share White and Military			-0.190^{***}	-0.211^{***}
			(0.066)	(0.066)
White x Cohort Share Minority, Military				-0.036
				(0.025)
White x Cohort Share White, Military				0.044^{*}
				(0.023)
Controls	Full	Full	Full	Full
R^2	0.430	0.431	0.433	0.434
Num. obs.	2336	2336	2336	2336

Note: Table displays the effect of cohort composition on full sample officers' propensities to make arrest of Blacks for low-level crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3) with additional variables of interest listed in the table. such as addition of terms for cohort shares interacted with an officer being white. High edu refers to having a Bachelors degree or above before the academy. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.6: Effect of Cohort Composition on Arrest Propensity - Full Sample Excluding 2011 Cohort

	Shrunken Black Arrest Propensity							
	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort Share Black	-0.212^{**}	0.036^{***}	-0.120^{*}	0.026^{***}				
	(0.086)	(0.012)	(0.067)	(0.006)				
Cohort Share Non-Black Minority	-0.121^{**}	0.009	-0.043	0.005				
	(0.055)	(0.009)	(0.047)	(0.006)				
Cohort Share Minority, Start Age <27					-0.029	0.006	-0.117	-0.006
					(0.049)	(0.009)	(0.088)	(0.027)
Cohort Share Minority, Start Age [27, 32]					-0.093	0.005	-0.159^{*}	-0.002
					(0.069)	(0.007)	(0.093)	(0.017)
Cohort Share Minority, Start Age >32					-0.151^{***}	0.019^{***}	-0.187^{***}	0.012
					(0.041)	(0.007)	(0.050)	(0.013)
Controls	Full	Full	Full	Full	Full	Full	Full	Full
Sample	Exams 2006, 2010	Exams 2006, 2010	Full	Full	Full	Full	Exams 2006, 2010	Exams 2006, 2010
R^2	0.261	0.123	0.437	0.158	0.440	0.157	0.261	0.121
Num. obs.	1181	1181	2296	2296	2296	2296	1181	1181

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to arrest Blacks for low-level and serious crimes in even and odd columns, respectively, excluding the 2011 cohort which may have mixed 2010 exam-takers and 2006 exam-takers. The propensity is captured by officers' fixed effects using equation (2) and shrunken as described in Appendix A.6. The parameter estimates are based on the specification in equation (3). Full controls refers to the additional controls in Column (13) of Table 3 with exam fixed effects and controls for cohort shares of whites who are start between 27 and 32 and above 32. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

C Appendix C - Data

The data used in this study were obtained via FOIA request, in collaboration with the Invisible Institute and Chicago Data Collaborative, and generously shared by Rachel Ryley.

Demographics Data on officer demographics were obtained via multiple FOIA request to the Chicago Police Department. These data include information on officers extending as far back as the 1940's to the present (2021). The core demographic data includes name, race (ethnicity), start date, resignation date, and gender. Additional data sets relating to officer's language abilities were obtained for more recent officers (i.e., those in the data for this study), which were used to determine if the officer reported being able to speak Spanish. Similarly, whether or not an officer was in the military was also obtained for the present set of officers. Educational attainment records were also obtained indicating where, when, and what degree (if any) was obtained by each officer– this data is much less complete than other data sets but is most complete for officers starting around the main sample. For simplicity, educational data was summarized for this study as an indicator ("high edu") if the officer had reported obtaining a Bachelors degree or higher (e.g., masters, law degree, doctorate) before they started at the academy.

Salary Salary data, obtained via FOIA to the Department of Human Resources, contains salary, pay grade (rank), and promotion information for officers between 2002 and 2020. This data is important as it allows us to focus on 'regular' police officers, i.e., D1 employees, and filter out promoted employees (sergeants, detectives, etc.). Importantly, this data contains officers' age at hire, allowing for very close approximation of their actual birth date and thus their exact age upon starting at the academy.

Unit History Officers' official unit assignments were obtained via FOIA to the CPD. This data indicates the dates on which an officer began and ended their official assignment to a specific unit.

Daily Assignments On a day to day basis, officers work specific beat assignments (alphanumeric codes that relate to function and location), are on specific watches, are or are not present for duty, are absent for some reason, are assigned to specific cars, and work between specific times. This information is contained within the daily assignment data, referred to in the text often as "AA" data. This data was obtained for the 22 (25 pre-2013) geographic units focused on in this study via FOIA request (for years 2010-2011 and 2016-2018) and shared by Rachel Ryley (for 2012-2015). Additional information on officer 'roles' were obtained via FOIA request to the CPD which gave descriptions of almost all beat assignment code to clarify their meaning.

Trainings A training data set, supplementary data set to the AA data, was obtained via FOIA request covering the period of the study. Specifically, this contains the name and start time of classes/trainings officers attended. This is particularly useful for identifying which officers were consistently trained together during the academy within their cohorts.

Arrests Data on adult arrests in Chicago were obtained via FOIA request to the CPD. This data includes arrestee information (race, age, gender), identifying officer information, arrest date and time, crime type and description, and the officer's arrest role (primary, secondary, or assisting). The arrest severity (Serious or Low-Level) is by crime type. Serious crimes include all violent and property index crimes, non-index property, and non-index violent crime (such as domestic violence and all forms of sexual assault). See *Crime* for crime code information. All other crimes (e.g., traffic, gambling, prostitution, drug) are considered low-level.

Court Court data from the Circuit Court of Cook County was obtained through collaboration with the Invisible Institute and Chicago Data Collaborative. This data is used to link specific arrests to cases and thus court outcomes (i.e., guilty finding, dropped case, etc.). It contains cases through 2019.

Population Information on district populations is obtained from the 2010 US Census. Median income is based on 2014 ACS estimates.

Crime Raw crime data is obtained from the Chicago Data Portal, downloaded in August of 2020. Crime is classified based on FBI codes into violent, property, and other crime. Violence-related crime FBI codes are 1A/B (homicide/manslaughter), 2 (criminal sexual assault / rape), 3 (robbery), 4A/B (aggravated assault/battery), 8A/B (simple assault/battery). Property-related crime FBI codes are 5 (burglary), 6 (theft), 7 (motor vehicle theft), 9 (arson), 10-13 (deceptive practices/fraud/stolen property), 14 (criminal damage). Index crime codes are 1A, 2, 3, 4A, 4B, 5, 6, 7, 9. All other crimes are classified as other and non-index, e.g., prostitution, gambling, trespassing, narcotics. Arrest data have the same classifications using FBI codes.