Learning by Doing in Law Enforcement

Jeremy West*

UC Santa Cruz

November 2018

Abstract

This study examines task-specific learning by police officers. Using data on twenty million traffic stops, I show that rookie officers become significantly more likely to find contraband as they accumulate experience, despite searching the same fraction of stopped vehicles. This indicates that officers learn to more accurately select which drivers to search. I find these productivity gains accrue primarily from changes to search behavior within rather than across driver demographic groups, supporting an explanation of skill improvements instead of increasing statistical discrimination. This evidence suggests that, while not eliminating racial bias, policing experience partially corrects for inefficiencies from stereotyping.

JEL: D24, D91, J18, K42

Keywords: Police stops, personnel experience, demographic biases

*Economics Department, University of California at Santa Cruz, and The E2e Project: westj@ucsc.edu. I thank the Stanford Open Policing Project for sharing the data used in this study. This work additionally benefit from valuable comments by George Bulman, Jesse Cunha, Carlos Dobkin, Robert Fairlie, Laura Giuliano, Benjamin Hansen, Mark Hoekstra, Vijetha Koppa, Natalia Lazzati, Justin Marion, Jonathan Meer, and Paul Oyer. Any errors in this manuscript are my own. I have no material financial interests related to the research described in this paper and provide all data and code used in this study at https://people.ucsc.edu/~jwest1/.

1 Introduction

At the turn of the nineteenth century, psychologists William Bryan and Noble Harter (1899) demonstrated that the productivity of telegraph operators improves significantly with practice. Since this pioneering work, researchers have explored how occupational experience forms human capital in settings ranging from subsistence farming to Uber driving (Foster and Rosenzweig, 1995; Cook, Diamond, Hall, List, and Oyer, 2018). Even among academic scholars, "a byproduct of producing research is learning" (Levin and Stephan, 1991). Broadly, two findings emerge from this literature. The first is that "task-specific human capital" is at least as important to worker productivity as is general knowledge and training (Gibbons and Waldman, 2004; Gathmann and Schönberg, 2010; Ost, 2014). The second is that, provided that workers are given frequent feedback regarding their performance, occupational learning curves tend to be steep initially before soon leveling off (e.g. Shaw and Lazear, 2008; Harris and Sass, 2011; Levitt, List, and Syverson, 2013; Haggag, McManus, and Paci, 2017).

In this study, I examine the productivity of police officers as they make contraband search decisions during traffic stops. This is a compelling context in which to evaluate learning by doing for several reasons. Most importantly, these decisions carry significant economic welfare consequence. The U.S. Bureau of Justice Statistics estimates that police officers find contraband in only 8.4 percent of the 870,000 traffic stop searches conducted nationally, equating to nearly 800,000 failed searches each year (Eith and Durose, 2011). Beyond the direct inefficiency from this misallocation of policing resources, fruitless searches create substantial time and stress cost for innocent civilians, and they invoke significant equity concerns regarding disparate treatment of racial and ethnic minorities (Engel and Calnon, 2004; Gelman, Fagan, and Kiss, 2007; Pierson, Simoiu, Overgoor, Corbett-Davies, Ramachandran, Phillips, and Goel, 2017). A second feature of the setting is that police officers' search productivity is directly observable, in contrast to the indirect measures of performance such as wages or teachers' "value added" to student achievement that are typically examined in studies of occupational learning (e.g. Murnane and Phillips, 1981; Dustmann and Meghir, 2005; Lazear, 2006; Gathmann and Schönberg, 2010; Harris and Sass, 2011). Finally, because police officers typically find contraband only several times per year, their traffic search efficiency can yield insights about task-specific learning when workers are provided with sparse feedback from which to learn.

Building on Becker's (1968) longstanding framework for policing strategy, I construct a simple theoretical model of officer learning in traffic stops. Conceptually, my model centers around an officer using information gleaned from interacting with previously stopped drivers to update, à la Bayes and Price (1763), how the officer forms beliefs about whether a currently stopped driver is carrying contraband. As an officer accumulates traffic stop experience, learning by doing implies that the officer's accuracy at identifying the presence of contraband should improve. The empirically testable implication is that, for a given (opportunity) cost of conducting a search and a given propensity for a stopped driver to be carrying contraband, an officer's "hit-rate" likelihood of finding contraband during a search will increase with experience. My model then provides a method to determine whether the mechanism(s) underlying any learning is officer skill improvement or increasing use of statistical discrimination. Specifically, the empirical test is whether the composition of searched drivers changes primarily within or across driver attributes such as demographic categorization that could be used to facilitate increased statistical discrimination.

Rather than focusing on a single police department, I empirically test the model's predictions using data on police stops during 2001-2016 across eight states, which comprise over twenty percent of the total U.S. population and span all four Census Regions. The data, obtained via public records requests made by Pierson et al. (2017), include nearly twenty million traffic stops made by about 5000 involved officers who are first observed as a rookie. Methodologically, my empirical identification uses within-officer variation, examining how search behavior and productivity change as an officer accumulates stop experience. These officers' average rookie hit-rate is only 21 percent, with significant heterogeneity in hit-rates across drivers by race and gender, providing substantial scope for productivity improvements both from skill development and from increased statistical discrimination.

I find robust evidence of task-specific learning in traffic stops. Consistent with the prediction of the model, I show that rookie officers significantly increase their productivity at finding contraband as they accumulate experience. Quantitatively, the average rookie will obtain a 56 percent improvement (11.74 percentage points) in contraband search hit-rate with ten years of experience. This finding is not attributable to officers' changing search frequencies or the composition of stopped drivers. Moreover, it is invariant to controlling for the composition of when an officer makes stops within the week or calendar year, as well as for broader nonlinear time trends. I also rule out explanations based on survivor bias, such as if police supervisors systematically deployed officers who are innately more successful at finding contraband to conduct higher volumes of traffic stops. Instead, I find that the primary mechanism underlying officers' search productivity improvements is changes to the search composition within driver demographic groups, supporting that the learning occurs predominantly via skill improvements rather than increased use of statistical discrimination.

This study makes several contributions. Most directly, my findings illustrate that task-specific human capital partially corrects for inefficiencies from stereotyping in law enforcement. A growing literature shows that a significant share of the variation in arrests and other treatment by police is attributable to heterogeneity in individual officers' behavior (Goncalves and Mello, 2017; Bulman, 2018; Weisburst, 2018; West, 2018). In recent closely-related work, DeAngelo and Owens (2017) demonstrate that more experienced state troopers tend to issue more frequent traffic citations, and concurrent work by Horrace, Jung, and Rohlin (2018) also finds that police contraband search productivity improves with officer experience. Neither of these studies, however, identifies whether the underlying mechanism for these productivity gains from experience is skill improvements or changing use of discrimination.

My findings also add to the discussion on reforming public sector compensation policies (Ebbinghaus, 2006). Because more experienced personnel add a larger wage burden to government budgets, policymakers increasingly offer early retirement incentives as a means of reducing labor costs (Fitzpatrick and Lovenheim, 2014). Using the *Texas Tribune*'s Government Salaries Explorer, I find that the wages of state police officers increase on average by about 1.82 percent in current dollars per year of experience.[1] Provided that my finding of ongoing productivity improvements generalizes to other policing responsibilities, this evidence supports that police officers are paid an efficiency wage for their service.

Finally, my study contributes to the larger literature on demographic bias in policing (e.g. Knowles, Persico, and Todd, 2001; Anwar and Fang, 2006; Antonovics and Knight, 2009; Fryer, 2016; West, 2018). Much of this literature focuses on determining whether observed disparities in police treatment result from statistical or preference-based discrimination. My evidence that police officers appear not to alter their use of statistical discrimination with experience provides insights, at least indirectly, to this significant literature. Given the large welfare consequences noted earlier from fruitless police searches, my findings also hold direct policy implications. For instance, increased deployment of more experienced officers to field work could yield substantial efficiency gains in law enforcement.

2 Model

Police officers make traffic stops for various reasons, ranging from observed violations (e.g. speeding) or suspicion of criminal activity, to reports of dangerous driving provided by other motorists (Engel and Calnon, 2004; Eith and Durose, 2011). Following the literature on

[1] I computed this average salary increase using cross-sectional data from https://salaries.texastribune.org.

motor vehicle searches, I take the sample of stopped cars as the population and do not model officers' stopping decisions (e.g. Knowles et al., 2001; Dharmapala and Ross, 2004; Anwar and Fang, 2006; Persico and Todd, 2006; Bjerk, 2007; Persico and Todd, 2008; Antonovics and Knight, 2009; Simoiu et al., 2017; Horrace et al., 2018).[2] After signaling to a driver to pull over and stop, the officer initiates a record of the stop using the police radio before personally interacting with the driver. Although "there is no such thing as a routine traffic stop," typically the officer informs the driver of the reason for the stop and records some administrative information such as driver license and insurance details.[3] Next, the officer may decide to conduct a vehicle search in order to seize any contraband such as stolen goods or illicit drugs or weapons.[4] I model this search process.

In the model, an officer's search decisions result from a constrained optimization problem. I model the officer's objective as maximizing the total amount of contraband found, subject to a limit on the officer's total volume of searches. This resource constraint could be due to external factors such as restrictions imposed by supervisors, or simply a limit to the officer's available time.[5] Formally, let $d_i \in \{0,1\}$ be an indicator for whether driver i is carrying contraband and $\tilde{d}_i \in [0,1]$ denote the officer's belief about the probability that $d_i = 1$ when making the decision of whether or not to conduct a search, $s_i \in \{0,1\}$. For some given volume of stops $N \in \mathbb{Z}^+$ and search rate limit $R \in (0,1)$, the officer makes search decisions to solve:

$$\underset{s_i \in \{0,1\}}{\text{Max}} \sum_{i=1}^{N} s_i \tilde{d}_i \qquad s.t. \qquad \frac{\sum_i s_i}{N} < R$$

From the officer's perspective, however, the problem is sequential rather than simultaneous. In a particular traffic stop, the officer's search decision solves: $\text{Max}_{s_i \in \{0,1\}} \, s_i \cdot (\tilde{d}_i - r)$. Here, the search cost r can be interpreted as the officer's opportunity cost of searching. To avoid trivial cases, I assume $r \in (0,1)$, so that the officer will search driver i if and only if $\tilde{d}_i > r$.

Next, I incorporate the potential for learning into the model. One clear implication of learning is that an officer's accuracy at predicting the presence of contraband should improve

[2] Of course, the composition of drivers that an officer stops could be a dimension of learning, I empirically explore this possibility and find evidence refuting this as a mechanism for search productivity improvements.

[3] c.f. http://www.policemag.com/channel/careers-training/articles/2012/11/traffic-stops.aspx

[4] Although many police searches require a search warrant, an exception for motor vehicles was established by the U.S. Supreme Court in *Carroll v. United States* (1925).

[5] Typical models in the literature do not impose search volume constraints, instead assuming that "officers can search all drivers if they so choose" (Antonovics and Knight, 2009). My assumption of binding search volume constraints is supported by the empirical findings.

with experience. From the analyst's perspective, learning implies that the expected prediction error for an officer should decrease in magnitude as the officer's stop experience increases: $\mathbb{E}\Big[\ |\ d_i - \tilde{d}_i\ |\ \Big] < \mathbb{E}\Big[\ |\ d_{i-1} - \tilde{d}_{i-1}\ |\ \Big]\ \ \forall i$, in which the driver subscript sequentially indexes stops made by the same officer. As officers' internal beliefs $\tilde{d}_i$ are unobservable, to facilitate empirical tests for learning I assume that each officer faces a constant opportunity cost of searching over time, i.e. that r is constant within-officer.[6] Given this assumption of constant search costs, for which I provide empirical support below in Section 4, a straightforward empirical test for learning is to evaluate whether an officer's rate of finding contraband (per stop or per search) increases with stop experience. That is, I test whether unconditionally $\mathbb{E}\Big[d_i\Big] > \mathbb{E}\Big[d_{i-1}\Big]\ \ \forall i$ and whether $\mathbb{E}\Big[d_i \mid s_i = 1\Big] > \mathbb{E}\Big[d_{i-1} \mid s_{i-1} = 1\Big]\ \ \forall i$.[7]

The model described so far provides testable predictions for whether officers learn to more accurately select which drivers to search. I next focus on identifying the underlying mechanism(s) for any search productivity improvements. In particular, I evaluate whether the improvements occur through increasing officer skill or increasing use of statistical discrimination. If present, the former would undoubtedly be a form of learning by doing, whereas the latter need not involve learning *by doing*. To distinguish between these potential mechanisms, let the officer's prediction $\tilde{d}_i = f(Z_i; \tilde{\alpha})$, in which Z_i is a vector of attributes particular to the driver and $\tilde{\alpha}$ is an officer-specific vector of parameters that the officer uses to map these driver characteristics into the contraband prediction $\tilde{d}_i$. Plugging $f(Z_i; \tilde{\alpha})$ into the definition of learning to reduce prediction error,

$$\mathbb{E}\Big[\ |\ d_i - f(Z_i; \tilde{\alpha})\ |\ \Big] < \mathbb{E}\Big[\ |\ d_{i-1} - f(Z_{i-1}; \tilde{\alpha})\ |\ \Big]\ \ \forall i.$$

It is clear from this formulation that at least some of the elements of $\tilde{\alpha}$ must be converging towards the true parameters $\alpha^\star$ as the officer accrues stop experience. Note that Z_i includes all information available to the officer at the time of the search decision: descriptive attributes such as the driver's demographics and the age of the vehicle, but also behavioral factors such as the driver's levels of nervousness, aggression, and evasiveness. While some components of Z_i such as demographic categorization can support certain search patterns based on statistical discrimination, mapping factors such as driver nervousness into contraband propensity can be done only using the officer's own skill. Thus, a compelling empirical

[6] Although models in the literature on police searches typically allow for heterogeneity in search costs across officers (as do I), they similarly assume that search costs remain constant within-officer (e.g. Anwar and Fang, 2006; Persico and Todd, 2006; Antonovics and Knight, 2009; Horrace et al., 2018).

[7] An implicit assumption in these tests is that there is no change over time in stopped drivers' propensity to be carrying contraband. Empirically, I can control for this possibility using time period fixed effects.

test for the mechanism(s) underlying learning is to evaluate whether the composition of searched drivers is changing primarily within or across components of Z_i that could be used to facilitate increased statistical discrimination.

3 Data

My empirical study employs a large panel of state police stops across eight states, sourcing data from Pierson et al. (2017). One advantage of this approach is generalizability. In addition to spanning all four Census Regions and time zones, the included states had a combined population of 68 million people in 2017, comprising 21 percent of the total U.S. population.[8] A second advantage is statistical power. Although few details about each traffic stop are needed to test the predictions derived in Section 2, the empirical tests require a large quantity of data. Identifying learning uses within-officer variation, so statistical power scales with the number of officers rather than with the number of observed stops. Moreover, because police officers locate contraband in only a small fraction of stops – about one in 200 stops in this study – a substantial number of observations per officer are also necessary to evaluate learning in contraband searches.

The data initially include the population of electronic records of traffic stops made by the state police in these eight states. The panel is unbalanced across officers, but ranges overall from 2001-2016. Because my study focuses on officer learning, I restrict my analysis to officers who are first observed as a "rookie," which I define to be an officer whose first observed stop occurs at least 365 days after the earliest observed stop made in the state.[9] After minor data cleaning, described in Appendix A, the analysis dataset includes 19.4 million traffic stops made by 4728 officers. All stops are made by only a single officer, and I include in this study only officers that conducted at least one search of a stopped driver.

Whereas the data include an average of more than 4000 traffic stop observations per officer, the level of detail is fairly limited. For each stop, I observe the unique officer identifier; the state, date, and time of day; the driver's gender and race; and binary indicators for whether a search was conducted and whether any contraband (drugs, weapons, or stolen property) was found. The data do not include any demographic information about the

[8] Appendix Figure A.1 shows a map of the included states: Arizona, Connecticut, North Carolina, South Carolina, Texas, Vermont, Washington, and Wisconsin. State population data are from the Census Bureau at https://www.census.gov/data/tables/2017/demo/popest/state-total.html.

[9] In principle, the study could also examine learning among more senior officers. However, the experience level for non-rookie officers is truncated and thus uncertain.

officers, nor do they indicate how much or what kind(s) of contraband was seized.[10]

Table 1 provides summary statistics on these traffic stops. The first two columns show the fraction of stops and searches, respectively, for which each indicator covariate equals one. Officers search a driver's vehicle for contraband in 2.14 percent of stops, and they find contraband in 0.53 percent of stops. This yields an average "hit-rate" search efficacy for contraband of 24.81 percent (in the data, officers never locate contraband when the record shows that no search was conducted). The remainder of Columns (1) and (2) shows the fraction of stops and searches that include a driver of a given demographic characteristic. As these statistics mainly reflect the composition of drivers on the road, a more informative comparison is provided in Columns (3) and (4), which respectively show the average search rate and (conditional on search) contraband finding hit-rate by driver demographic attribute.

These statistics reveal two intriguing insights. First, the significant variation in both search rates and hit-rates across drivers of differing demographic characteristics indicates that substantial efficiency gains might be expected from officers reallocating searches across driver groups. For instance, intensified statistical discrimination would support officers increasing search frequencies for both black and white drivers, while decreasing searches of Hispanics. More strikingly, even for the demographic groups that officers are most effective at searching, more than 70 percent of searches yield no contraband; overall, three out of four searches are fruitless.[11] This illustrates that (by far) the bigger potential for search efficiency gains is from changing search behavior within rather than across demographic groups. In the following section, I evaluate both potential channels for improvements in officer contraband search decisions, testing predictions derived above in Section 2. Altogether, Table 1 indicates there is substantial potential for policing productivity gains from officer learning by doing.

4 Results

In this section, I use the data described in Section 3 to test the predictions of the model presented in Section 2. After demonstrating that officers' contraband search efficacy increases significantly with experience, I evaluate the potential mechanisms underlying this productivity improvement. I find that the primary channel is through changing search behavior

[10] The detail of the stop records varies extensively. For example, Arizona, Texas, and Wisconsin do not report driver age, which is reported by the other five states included in this study. Even within-state, some records include the county or geocoordinates of the stop, whereas others do not.

[11] Notably, this average search productivity is three times the national average, which may be attributable to my focus only on state police (Eith and Durose, 2011). Other studies find similarly low average hit-rates (e.g. Anwar and Fang, 2006; Persico and Todd, 2006; Simoiu et al., 2017).

within rather than across driver demographic groups, supporting an explanation of officer skill improvements rather than increased statistical discrimination.

4.1 Vehicle search productivity improves with experience

Before testing the predictions of the model econometrically, I explore the basic relationship between officer experience and contraband search efficacy. Figure 1 plots the likelihood that an officer finds contraband in a vehicle search across different levels of accumulated experience. The experience term includes all stops made by the officer prior to that stop, regardless of whether a search was conducted. All included officers are initially rookies, and the same officers are shown across multiple levels of experience along the horizontal axis. The marker sizes indicate the number of officers observed at each experience level, using bins of 100 stops. In addition to plotting local averages, I plot a linear regression fit line, estimated using the underlying more granular microdata. This regression has an intercept of 21.12 percent and a slope of 1.081 percentage points per thousand stops, indicating that officers substantially improve their contraband hit-rates with experience.

This motivating visual evidence from the plot of the raw data strongly supports the presence of learning by doing. Next, I more formally examine the role of accumulated traffic stop experience using specifications of the following form.

$$\text{outcome}_{ij} = \text{experience}_{ij} \cdot \beta + \mu_j + \text{controls}'_{ij}\Gamma + \epsilon_{ij} \tag{1}$$

In Equation 1, outcome_{ij} is some outcome pertaining to the traffic stop of driver i made by officer j, such as whether a vehicle search was conducted; experience_{ij} is a linear term counting officer j's accumulated stop experience prior to stop i; μ_j is an officer fixed effect; and controls_{ij} include various fixed effects and covariates particular to the stop, such as the hour within the week and the date of the calendar year. Table 2 presents estimates using Equation 1. Each cell shows an estimate of β from a separate linear model, using different outcomes across the panels. Standard errors in parentheses are clustered by officer.

In Panel [A], I first evaluate how officers' monthly stop volume changes with experience. Across Columns (1)-(4), which incorporate various fixed effects (discussed below), the estimates indicate that each 1000 stops of prior experience is associated with the officer making about four fewer stops per month. In Column (5), which adds fixed effects for the 16 years of the sample, the estimated magnitude shrinks by about half, suggesting that some of this decline reflects a broader systematic reduction in traffic stop volume (per officer), but there remains a strong negative relationship between officer experience and stop frequency. Given that officers' average monthly volume is about 84 stops (just over 1000 per year), these estimates equate to about a two to five percent reduction in traffic stop volume per year of police experience. Although the model presented in Section 2 neither assumes nor predicts a relationship between officer experience and stop frequency, this finding is not surprising. State police officers have numerous responsibilities, and it seems quite plausible that more senior officers could be relatively more likely to be assigned to other tasks such as vehicle crash investigations or sobriety checkpoints. However, this finding has two important implications. First, any learning by doing occurs despite officers having reduced opportunity to practice as they accumulate experience. Second, the most experienced officers – who are more productive at finding contraband – are also less likely to be making traffic stops.

In Panel [B] of Table 2, I test the key assumption of the model that each officer faces a constant search cost over time. Specifically, this panel assesses how the likelihood of conducting a vehicle search varies with an officer's stop experience. The evidence is clear that officers do not systematically change their search rates of stopped drivers as they accumulate experience. Across all columns, the estimates are statistically and economically insignificant. If officers' search costs were declining over time – or if the reduction in stop frequency shown in Panel [A] was due to officers systematically stopping fewer drivers who have high search costs or low expected contraband propensity – then officers would be increasing their search rates as they accrue experience. The null effects for changing search rates are consistent with officers' facing constant search costs despite their level of experience, as is assumed in the model in Section 2, as well as by models in the literature more broadly (e.g. Anwar and Fang, 2006; Persico and Todd, 2006; Antonovics and Knight, 2009; Horrace et al., 2018).

Panel [C] shows tests for learning, presenting estimates for the likelihood that an officer finds contraband during a stop in reduced-form, unconditional on whether a search is conducted. In Column (1), which includes officer fixed effects and no controls, the estimated coefficient indicates that officers are 0.022 percentage points (s.e.= 0.003) more likely to find contraband in a stop for each thousand stops of prior experience they have accumulated. On a mean contraband-finding rate of 0.53 percent, this estimate is economically significant, especially when considering that officers on average make about 1000 stops per year. The estimate is barely changed in Column (2), which adds fixed effects for the 168 hours of the week, or in Column (3), which also adds fixed effects for the 366 dates of the calendar year.

The final two columns of Table 2 address alternative explanations. One consideration is that, rather than learning, more experienced officers may be assigned more often to time shifts that have systematically better opportunities to find contraband. Indirectly, this explanation is largely ruled out by the null effects for search rates. As further evidence, Column (4) uses officer by hour-of-week fixed effects, assessing how contraband finding efficacy varies within the same officer during the same hour of the week. The estimated coefficient of 0.024 percentage points (s.e.= 0.003) is identical to that in the previous columns.[12] Finally, I evaluate whether these findings simply capture some broader change(s) in drivers' propensity to carry contraband. The estimate in Column (5) remains unchanged as I add fixed effects for the sixteen years of the sample, using within-officer by hour-of-week and within-year variation, while controlling for the calendar date.

Given the above results, it is unsurprising in Panel [D] of Table 2 to see that officers' search productivity significantly increases with stop experience. Quantitatively, each thousand stops' worth of experience is associated with about a 1.23 percentage points (s.e.= 0.091) improvement in an officer's expected hit-rate. On a rookie baseline hit-rate of 21.12 percent (from Figure 1), this predicts that the average rookie will obtain a 56 percent increase (11.74 percentage points) in the rate of finding contraband per search after ten years of experience – quite an economically significant improvement. Altogether, Table 2 provides ample evidence in support of the model of learning by doing in traffic stops. Having established that officer experience significantly impacts contraband search performance, I next focus on identifying the underlying mechanism(s) for these productivity improvements.

4.2 No evidence of survivor bias in traffic stop volumes

State police officers have numerous tasks, and the task assignment of officers is not arbitrary. Even before accruing much traffic stop experience, police officers differ substantially in their contraband search productivity. For officers during their rookie year, the 25th percentile hit-rate is zero percent whereas the 75th percentile is 34 percent. Some of this variation is attributable to heterogeneous search volumes; but, even restricting to officers who made at least ten searches during their rookie year, the 25th percentile hit-rate is seven percent and the 75th percentile is 31 percent. Thus, police supervisors have scope to assign officers who had comparatively better rookie contraband search productivity to conduct higher traffic stops volumes in later years. This would bias the tests presented earlier towards showing

[12] A related consideration is that more experienced officers may patrol locations with systematically better opportunities for finding contraband. State police officers conduct almost exclusively highway stops, but there could be selection onto "back roads" versus Interstates, for example. The spatial identifiers in the data are not consistently fine enough to directly address this potential mechanism, but such a practice is also largely ruled out by the null effects for changing search rates.

learning. I evaluate whether such survivor bias is a concern using tests of the following form.

$$\text{total\_stops}_j = \text{rookie\_stops}_j \cdot \phi + \text{rookie\_contraband}_j \cdot \delta + \text{controls}_j' \Lambda + \epsilon_j \tag{2}$$

In the specification of Equation 2, total_stops_j is the total number of observed stops in the data that are made by officer j; rookie_stops_j is the number of stops that the officer made during his or her rookie year; $\text{rookie\_contraband}_j$ is the total number of searches the officer conducted during the rookie year that yielded contraband; and controls_j are time-invariant officer covariates such as the officer's state and starting date. The coefficient of interest is δ, which would be large and positive if there were significant survivor bias.

Table 3 presents the results of these tests for survivor bias. Officers make on average 4106 observed stops in total (961 annualized), averaging 962 stops during their rookie year and 1788 stops during their first two years. Consistent with intuition about officer deployment to heterogeneous tasks, the estimates for ϕ are statistically and economically significant: officers who made more stops during their rookie year(s) are observed making more stops in total. Quantitatively, an officer who made 100 additional stops (about 10 percent) as a rookie will make about 292 more observed stops in total. The magnitude and statistical significance of this relationship continues to hold across Columns (2)-(4), which add a covariate for rookie contraband finding as well as fixed effects for the officer's starting date and state of employment. The estimated coefficient for ϕ is smaller in Column (5), which uses each officer's rookie 2-years, but this attenuation is a mechanical artifact of the predictor variable absorbing a larger share of the total stops.

In contrast to the strong predictive power of officers' rookie stop volumes, there is no relationship between officers' rookie search efficacy and total volume of traffic stops. Taking the point estimates for δ at face value, the coefficients suggest that, at most, an officer with *double* the average first-year contraband finding productivity will make 18 additional observed stops in total, 0.4 percent of the mean. This null effect could be explained by the relatively low information content of rookie search efficacy. Alternatively, it could reflect the multifaceted objectives of traffic stops, such as minimizing traffic fatalities in addition to seizing contraband (Bates et al., 2012; DeAngelo and Hansen, 2014). In any case, these tests convincingly rule out survivor bias as a mechanism for the productivity gains from experience demonstrated earlier.

4.3 Evidence supports mechanism of officer skill improvement

The findings presented so far show that police officers do not alter their overall search rates as they accrue traffic stop experience and that officers with higher baseline search productivity do not systematically make more stops. Thus, the demonstrated productivity gains associated with experience indicate that officers learn to more accurately select which drivers to search.[13] There are two broad mechanisms by which officers might select for a more accurate composition of drivers to search: they could increase their use of statistical discrimination, or they could improve their skill at evaluating drivers' idiosyncratic attributes and behavioral tells.

I test for both potential mechanisms in Table 4, presenting estimates of Equation 1 separately for various subsets of drivers based on selected demographic characteristics. Panel [A] evaluates how monthly stop frequencies change with officer experience, Panel [B] assesses the likelihood that an officer conducts a search during a stop, and Panel [C] examines the likelihood that the officer finds contraband during the stop. All specifications in the table use the same fixed effects as in Column (4) of Table 2, and the officer experience term is defined using all prior stop experience.

Broadly, Table 4 does not support that traffic stop experience substantially changes officers' use of statistical discrimination. Across Panel [A], drivers of all demographic groups are stopped less frequently by more experienced officers. As discussed above, this is more likely to be attributable to changing task assignments rather than intentional officer discretion in stop volumes. Consistent with the statistical guidance provided in Table 1 (and discussed near the end of Section 3), Panel [B] shows that officers decrease search frequencies of Hispanic drivers as they accrue experience. However, they also increase search rates of female drivers, in contrast to the pattern supported by heightened statistical discrimination. Accumulated experience is not associated with any significant changes to the search rates of male, black, or white drivers.[14] In Panel [C], by contrast, it is clear that significant search productivity improvements are present within all demographic groups. Across the columns of Panel [C], the increases in contraband-finding efficacy range from 2.9 to six percent of the

[13] An alternative explanation is that officers improve in the search procedure itself. Given the magnitude of the hit-rate improvements, it is difficult to hold that these productivity gains are attributable to officers learning to more thoroughly search vehicles. Moreover, such an explanation implies that officers should be increasing search rates as well.

[14] Officers might be increasing their use of statistical discrimination along other dimensions, such as based on the age of a driver's vehicle. However, that they do not substantially change their use of statistical discrimination along observable demographic driver attributes makes it less plausible that they are significantly changing their use of statistical discrimination along unobservable (to the analyst) dimensions.

respective subgroup means, and all of the point estimates lie within the 95 percent confidence interval for the full-sample estimate in Column (1), and vice versa.

More tellingly, comparing the estimates in Panel [B] to those in Panel [C] shows that the search productivity gains are much larger in magnitude than the changes to the demographic search composition. Proportionally, the biggest observed change in search rates (for Hispanics) is 1.5 percent in magnitude per 1000 stops of experience, whereas the increase in contraband finding is 4.5 percent of the mean for the same demographic group. Even more directly, the results show that, with increased experience, officers both significantly decrease the search rate of Hispanics *per stop* while significantly increasing contraband finding rates *per stop* among the same group. Collectively, Table 4 illustrates that the productivity gains from experience are primarily attributable to officers changing search behavior within demographic groups, supporting a mechanism of task-specific skill improvements rather than intensified statistical discrimination.

5 Conclusions

This study examines task-specific human capital development by police officers. Layering a Bayesian updating framework onto a model of officer policing strategy, I develop several tests for learning by doing in contraband searches, which I then empirically evaluate using data on nearly twenty million traffic stops made by about 5000 police officers.

Consistent with the model, I find that police officers' productivity in searching for contraband significantly increases as they accumulate stop experience. Supporting exercises show that this finding is not being driven by selection bias, nor by any obvious channels attributable to police agencies or broader secular patterns. Instead, I find that the primary mechanism is officers' choices about the composition of searched drivers within demographic categories, supporting that learning occurs predominantly through skill improvements.

This study yields several contributions. Most directly, my findings show that task-specific human capital partially corrects for inefficiencies from stereotyping in law enforcement. In addition, by providing evidence that policing experience appears not to alter officers' use of statistical discrimination, my study contributes to the broader literature on demographic bias in policing. Finally, my findings hold direct policy implications. Given the large welfare consequences and social justice concerns associated with fruitless police searches, the evidence from this study supports that policies to revise the training and task assignments of police officers can substantially improve the efficiency of law enforcement.

References

K. Antonovics and B. G. Knight. A new look at racial profiling: Evidence from the Boston police department. *The Review of Economics and Statistics*, 91(1):163–177, February 2009.

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

L. Bates, D. Soole, and B. Watson. The effectiveness of traffic policing in reducing traffic crashes. In T. Prenzler, editor, *Policing and Security in Practice*, chapter 6, pages 90–109. Palgrave Macmillan, 2012.

T. Bayes and R. Price. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, pages 370–418, January 1763.

G. S. Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, March-April 1968.

D. Bjerk. Racial profiling, statistical discrimination, and the effect of a colorblind policy on the crime rate. *Journal of Public Economic Theory*, 9(3):521–545, June 2007.

W. L. Bryan and N. Harter. Studies on the telegraphic language: The acquisition of a hierarchy of habits. *The Psychology Review*, 6(4):345–375, July 1899.

G. Bulman. Law enforcement leaders and the racial composition of arrests: Evidence from overlapping jurisdictions. UC Santa Cruz working paper, May 2018.

C. Cook, R. Diamond, J. Hall, J. A. List, and P. Oyer. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Stanford University working paper, June 2018.

G. DeAngelo and B. Hansen. Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy*, 6(2):231–257, May 2014.

G. DeAngelo and E. G. Owens. Learning the ropes: General experience, task-specific experience, and the output of police officers. *Journal of Economic Behavior & Organization*, 142:368–377, October 2017.

D. Dharmapala and S. L. Ross. Racial bias in motor vehicle searches: Additional theory and evidence. *Contributions in Economic Analysis & Policy*, 3(1), September 2004.

C. Dustmann and C. Meghir. Wages, experience, and seniority. *Review of Economic Studies*, 72(1):77–108, January 2005.

B. Ebbinghaus. *Reforming Early Retirement in Europe, Japan and the USA*. Oxford University Press, July 2006.

C. Eith and M. R. Durose. Contacts between police and the public. Special Report NCJ 234599, U.S. Department of Justice, Bureau of Justice Statistics, October 2011.

R. S. Engel and J. M. Calnon. Examining the influence of drivers' characteristics during traffic stops with police: Results from a national survey. *Justice Quarterly*, 21(1):49–90, March 2004.

M. D. Fitzpatrick and M. F. Lovenheim. Early retirement incentives and student achievement. *American Economic Journal: Economic Policy*, 6(3):120–154, August 2014.

A. D. Foster and M. R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209, December 1995.

R. G. Fryer. An empirical analysis of racial differences in police use of force. NBER Working Paper No. w22399, July 2016.

C. Gathmann and U. Schönberg. How general is human capital? A task-based approach. *Journal of Labor Economics*, 28(1):1–49, January 2010.

A. Gelman, J. Fagan, and A. Kiss. An analysis of the New York City Police Department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, 102(479):813–823, September 2007.

R. Gibbons and M. Waldman. Task-specific human capital. *The American Economic Review: Papers & Proceedings*, 94(2):203–207, May 2004.

F. Goncalves and S. Mello. A few bad apples? Racial bias in policing. Princeton University working paper 608, March 2017.

K. Haggag, B. McManus, and G. Paci. Learning by driving: Productivity improvements by New York City taxi drivers. *American Economic Journal: Applied Economics*, 9(1):70–95, January 2017.

D. N. Harris and T. R. Sass. Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7-8):798–812, August 2011.

W. C. Horrace, H. Jung, and S. M. Rohlin. Police officer experience, learning and bias in traffic stops. Syracuse University working paper, January 2018.

J. Knowles, N. Persico, and P. Todd. Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, 109(1):203–229, February 2001.

E. P. Lazear. Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, 121(3):1029–1061, August 2006.

S. G. Levin and P. E. Stephan. Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 81(1):114–132, March 1991.

S. D. Levitt, J. A. List, and C. Syverson. Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy*, 121(4):643–681, August 2013.

R. J. Murnane and B. R. Phillips. Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review*, 1(4):453–465, Fall 1981.

B. Ost. How do teachers improve? the relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, 6(2):127–151, April 2014.

N. Persico and P. Todd. Generalising the hit rates test for racial bias in law enforcement, with an application to vehicle searches in Wichita. *Economic Journal*, 116(515):F351–F367, November 2006.

N. Persico and P. Todd. The hit rates test for racial bias in motor-vehicle searches. *Justice Quarterly*, 25(1):37–53, March 2008.

E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, V. Ramachandran, C. Phillips, and S. Goel. A large-scale analysis of racial disparities in police stops across the United States. Stanford University working paper, June 2017.

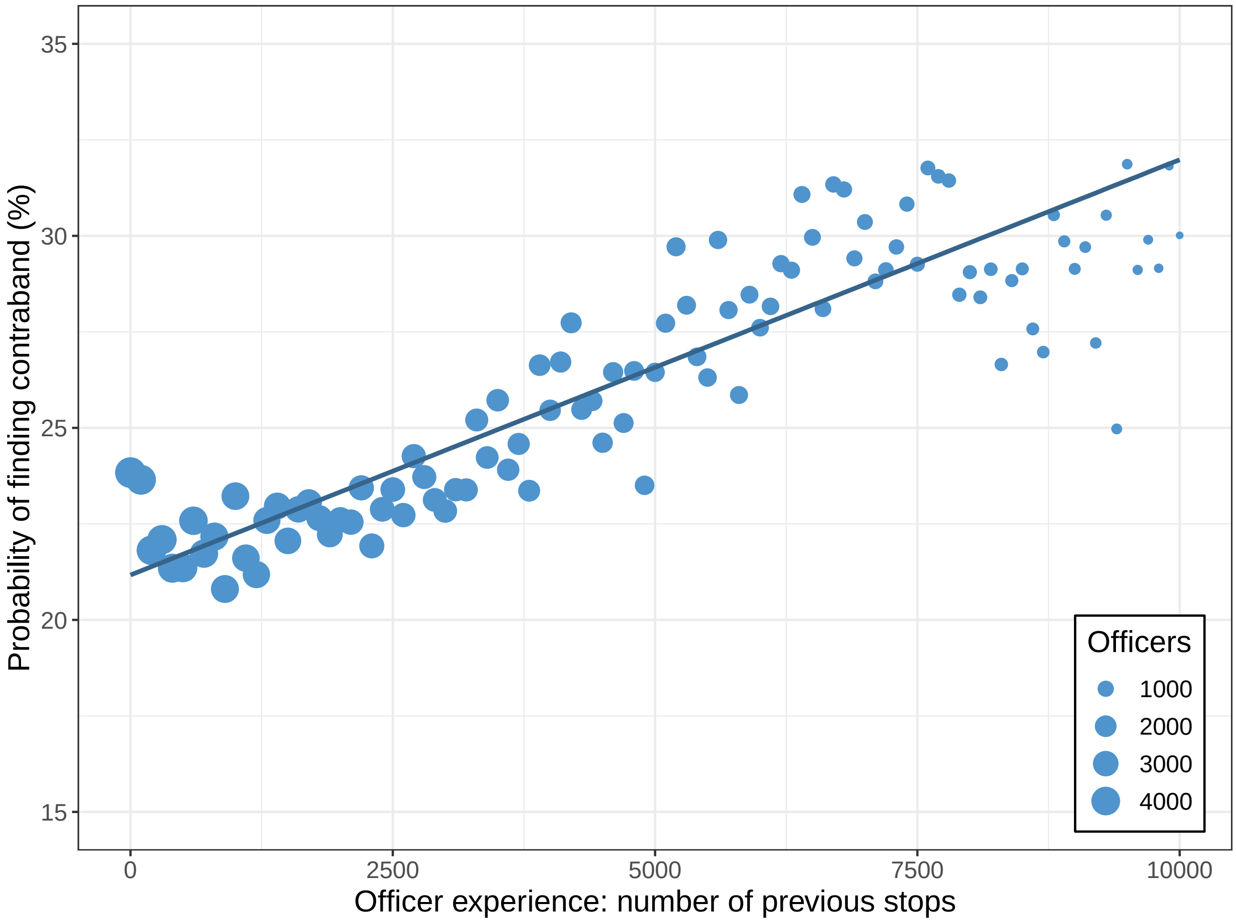
K. Shaw and E. P. Lazear. Tenure and output. *Labour Economics*, 15(4):705–724, August 2008.

C. Simoiu, S. Corbett-Davies, and S. Goel. The problem of inframarginality in outcome tests for discrimination. *Annals of Applied Statistics*, 11(3):1193–1216, September 2017.

E. Weisburst. "Whose help is on the way?" The importance of individual police officers in law enforcement outcomes. UCLA working paper, September 2018.

J. West. Racial bias in police investigations. UC Santa Cruz working paper, October 2018.

Figure 1: Likelihood of finding contraband during a search by officer experience



Notes: Figure 1 includes stops by state police in Arizona, Connecticut, North Carolina, South Carolina, Texas, Vermont, Washington, and Wisconsin. All included officers are initially rookies and the same officers are included across multiple levels of experience along the x-axis. The marker sizes indicate the number of officers observed at each experience level. The probability of finding contraband (search hit-rate) is defined as the fraction of searches for which any contraband was found, based on a binary outcome for each search conducted in the underlying microdata. The officer experience term counts all stops made by the officer prior to that stop, regardless of whether a search was conducted. For visual clarity, the markers are binned into hundreds of stops experience and truncated at 10,000 stops, the 95th percentile within sample. The linear regression fit line is estimated using the underlying microdata and has a y-intercept of 21.12 percent and a slope of 1.081 percentage points per thousand stops.

Table 1: Summary statistics on state police stops

	Fraction of		Average	
	Stops (1)	Searches (2)	Search rate (3)	Hit rate (4)
Search conducted	0.0214	1.0000		
Contraband found	0.0053	0.2481		
Female driver	0.3256	0.2048	0.0138	0.2382
Male driver	0.6744	0.7952	0.0259	0.2505
Asian driver	0.0170	0.0115	0.0148	0.1645
Black driver	0.1569	0.1852	0.0259	0.2877
Hispanic driver	0.2178	0.3075	0.0310	0.1961
White driver	0.5905	0.4653	0.0173	0.2699
Total count	19,411,192	415,103	415,103	102,982

Notes: Table 1 includes stops by state police in Arizona, Connecticut, North Carolina, South Carolina, Texas, Vermont, Washington, and Wisconsin. All included officers are initially rookies. The first two columns show the fraction of stops and searches for which each covariate in rows takes a logical value of one. The third and fourth columns show the average search rate and contraband finding hit rate by driver demographic characteristics.

Table 2: Officer stop behavior and search efficacy by experience

	Specification				
	(1)	(2)	(3)	(4)	(5)
Panel [A] Number of stops made by officer per month (mean = 84.46)					
Experience: stops/1000	-4.179^{***}	-4.193^{***}	-4.173^{***}	-4.105^{***}	-2.298^{***}
	(0.195)	(0.195)	(0.195)	(0.196)	(0.426)
Panel [B] Likelihood officer searches during a stop (mean = 2.14%)					
Experience: stops/1000	-0.005	0.001	0.001	-0.002	0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.017)
Panel [C] Likelihood officer finds contraband during a stop (mean = 0.53%)					
Experience: stops/1000	0.022^{***}	0.024^{***}	0.024^{***}	0.024^{***}	0.024^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)
Panel [D] Likelihood officer finds contraband during a search (mean = 24.81%)					
Experience: stops/1000	1.221^{***}	1.235^{***}	1.230^{***}	1.003^{***}	1.481^{***}
	(0.091)	(0.091)	(0.091)	(0.131)	(0.311)
Officer fixed effects	Yes	Yes	Yes	-	-
Hour-of-week fixed effects	No	Yes	Yes	-	-
Date-of-year fixed effects	No	No	Yes	Yes	Yes
Off. by hour-of-week FE	No	No	No	Yes	Yes
Year fixed effects	No	No	No	No	Yes
Number of officers	4,728	4,728	4,728	4,728	4,728
Number of searches	415,103	415,103	415,103	415,103	415,103
Number of stops	19,411,192	19,411,192	19,411,192	19,411,192	19,411,192

Notes: *p<0.1; **p<0.05; ***p<0.01; Standard errors are clustered by officer. Each cell presents results from a separate linear model. Panel [A] evaluates how many stops the officer conducted during the month. Panel [B] assesses whether a search was conducted during the stop. Panel [C] examines whether contraband was found during the stop in reduced-form, including all stops regardless of whether a search was conducted. Panel [D] examines whether contraband was found for the subset of stops in which a search was conducted. Reported coefficients are for the linear officer experience term, which counts all prior stops made by the officer. The officer fixed effects are unique to each officer. The hour-of-week fixed effects include 168 values for the hours of the week. The date-of-year fixed effects include 366 values for the dates of the calendar year. The officer by hour-of-week fixed effects are interactions unique to each officer by hour of the week. The year fixed effects include 16 values for the years included in the sample (2001-2016).

Table 3: Identification tests for survivor bias

	Total observed stops made by officer (mean = 4106)				
	(1)	(2)	(3)	(4)	(5)
Rookie year total stops	2.922***	2.917***	3.238***	3.234***	
[mean = 962]	(0.063)	(0.065)	(0.079)	(0.086)	
Rookie year total contraband		1.849	3.748	3.537	
[mean = 4.87]		(4.847)	(4.438)	(4.469)	
Rookie 2-years total stops					2.075***
[mean = 1788]					(0.037)
Rookie 2-years contraband					−0.418
[mean = 9.43]					(1.808)
Constant	1,294***	1,290***			
	(74.843)	(75.518)			
Officers' initial date fixed effects	No	No	Yes	Yes	Yes
State fixed effects	No	No	No	Yes	Yes
Observations (officers)	4,728	4,728	4,728	4,728	4,728

Notes: *p<0.1; **p<0.05; ***p<0.01; Each column presents results from a separate linear regression model for the total within-sample stops made by each officer. Columns (1)-(4) use measures of experience and contraband finding during each officer's first year. Column (5) uses measures of experience and contraband finding pooled over each officer's first two years. The officers' initial date fixed effects are values for the date of the first observed stop made by the officer. The state fixed effects are values for the eight states included in the sample.

Table 4: Officer stop behavior and search efficacy by experience for demographic groups

	Driver demographic characteristic					
	All (1)	Female (2)	Male (3)	Black (4)	Hispanic (5)	White (6)
Panel [A] Number of stops made by officer per month						
Experience: stops/1000	-4.105^{***}	-0.939^{***}	-3.148^{***}	-0.348^{***}	-0.367^{***}	-3.264^{***}
	(0.196)	(0.073)	(0.132)	(0.052)	(0.089)	(0.163)
Coefficient/mean	-0.0486	-0.0350	-0.0567	-0.0273	-0.0208	-0.0682
Panel [B] Likelihood officer searches during a stop						
Experience: stops/1000	-0.002	0.016^{***}	-0.003	-0.006	-0.047^{***}	0.009
	(0.008)	(0.006)	(0.009)	(0.015)	(0.017)	(0.007)
Coefficient/mean	-0.0009	0.0115	-0.0011	-0.0023	-0.0153	0.0050
Panel [C] Likelihood officer finds contraband during a stop						
Experience: stops/1000	0.024^{***}	0.020^{***}	0.028^{***}	0.022^{***}	0.028^{***}	0.022^{***}
	(0.003)	(0.002)	(0.004)	(0.006)	(0.005)	(0.003)
Coefficient/mean	0.0447	0.0595	0.0432	0.0289	0.0453	0.0468
Mean monthly stops	84.46	26.79	55.49	12.73	17.66	47.89
Mean search rate	2.14%	1.38%	2.59%	2.59%	3.10%	1.73%
Mean contraband rate	0.53%	0.33%	0.65%	0.75%	0.61%	0.47%
Date-of-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Off. by hour-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of officers	4,728	4,636	4,717	4,604	4,601	4,680
Number of searches	415,103	84,982	330,022	75,861	125,969	190,613
Number of stops	19,411,192	6,157,957	12,753,618	2,925,028	4,058,837	11,006,341

Notes: *p<0.1; **p<0.05; ***p<0.01; Standard errors are clustered by officer. Each cell presents results from a separate linear model. Panel [A] evaluates how many stops the officer conducted during the month of drivers of selected demographic characteristics. Panel [B] assesses whether a search was conducted during the stop. Panel [C] examines whether contraband was found during the stop in reduced-form, including all stops regardless of whether a search was conducted. In Panels [B] and [C], the columns use the respective subsets of data corresponding to drivers of demographic characteristics indicated by the column titles. As in Table 2, reported coefficients are for the linear officer experience term, which counts all prior stops made by the officer. The date-of-year fixed effects include 366 values for the dates of the calendar year. The officer by hour-of-week fixed effects are interactions unique to each officer by hour of the week.

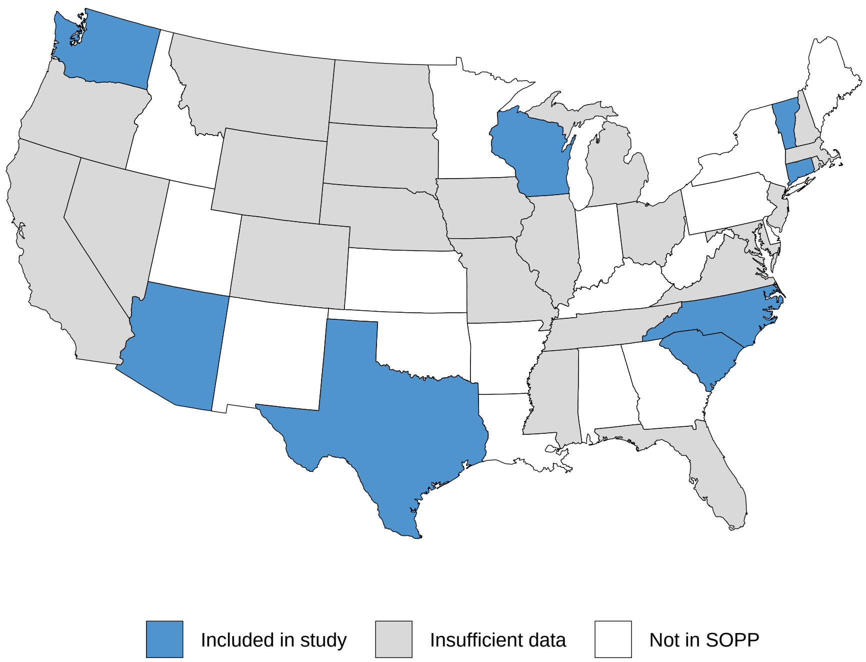
A Data Appendix

All data included in this study were originally provided by state agencies in response to public records requests made by the Stanford Open Policing Project (SOPP, openpolicing.stanford.edu). Researchers who are interested in replication or extension should reference Pierson et al. (2017) and the SOPP data ReadMe in addition to the data documentation that I provide for this study. Data availability varies substantially across states. Of the thirty-one states in the SOPP data, sixteen include officer identifiers. Of these sixteen states, eight do not consistently report officer search behavior or contraband finding. The remaining eight states are included in this study. State categorizations are mapped in Appendix Figure A.1.

The data as I received them were already partially cleaned by SOPP. I conducted moderate additional data cleaning and standardization across states, particularly for officer identifiers. This data preparation was done using my provided *R* code, and associated explanation and tabulations are provided in my ReadMe document. In particular, I correct for data entry heterogeneity such as leading-zero padding in the officer badge numbers and variation in the entry of sub-agency police divisional identifiers. In addition, although this study only focuses on officer contraband search behavior, the data and my associated cleaning/standardization include several additional police stop covariates and outcome variables. Again, data availability and specificity vary substantially across and within states.

After cleaning and standardizing officer identifiers, all stops are made by a single officer, a very minor restriction. I define a “rookie” officer to be a badge ID whose first observed stop occurs at least 365 days after the earliest observed stop made in the state. I drop observations on 6162 rookie officers who never conduct a single search in the data (many of these are likely typos in the entry of officers’ badge numbers). The resulting data examined in this study include 19,411,192 traffic stop observations for 4728 rookie officers who conducted at least one search.

Figure A.1: State data availability and inclusion



Notes: Figure A.1 illustrates state inclusion in the Stanford Open Policing Project (SOPP) data and in this study. There are thirty-one states included in the original SOPP data. Of these, fifteen states are excluded from this study because they do not include officer identifiers: California, Colorado, Illinois, Massachusetts, Maryland, Missouri, Montana, North Dakota, Nebraska, New Hampshire, Nevada, Oregon, Rhode Island, South Dakota, and Tennessee. An additional eight states are excluded because they do not report search behavior or contraband information: Florida, Iowa, Michigan, Mississippi, New Jersey, Ohio, Virginia, and Wyoming. This yields eight states that are included in this study: Arizona, Connecticut, North Carolina, South Carolina, Texas, Vermont, Washington, and Wisconsin. The remaining 19 states (including Alaska and Hawaii) did not provide SOPP with any data.