Not So Black and White: Uncovering Racial Bias from Systematically Masked Police Reports

Elizabeth Luh

August 9, 2019

Abstract

Biased police officers may purposely mis-record, or mask, the race of citizens that they interact with in order to evade detection. Indeed, journalists uncovered widespread evidence of such masking among Texas Highway troopers from 2010 to 2015. I propose a new test of racial bias in the presence of masking that is more powerful than standard tests and is well-suited to explore the rich heterogeneity in bias. Using various data-driven techniques to detect masking, I estimate that 24% of 130,240 searches were masked, with over half being Hispanic drivers being mis-recorded as white when searches failed to turn up contraband. I find that Hispanic and white troopers are biased against non-white motorists, with Hispanic motorists being treated the most unfairly. Using my model, I also find evidence of institutional racial bias and 'bad apple' troopers across Texas.

1 Introduction

In the United States, racial bias in the criminal justice system is a pervasive and prevalent issue. Disparities in the treatment of citizens of different races have been found in nearly all types of interactions between citizens and law enforcement, from motorist stops for black and Hispanic motorists (Harris (1999)) or in the more extreme cases of alleged excessive use of police force against minorities (Lind (2015))(for example: Michael Brown). These disparities carry over into each step of the justice system from airport screening (Persico and Todd (2005)), ticketing (Anbarci and Lee (2014), Goncalves and Mello (2017)), stop and frisk participation (Coviello and Persico (2013)), bail decisions (Arnold et al. (2018)), sentencing (Shayo and Zussman (2011)), parole (Anwar and Fang (2015)), and in capital punishment (Alesina and Ferrara (2014)). While researchers and policy makers have dedicated much focus to detecting and measuring racial bias, little attention has been given to the response of law enforcement officers to this heightened scrutiny.

This is a major concern since researchers can only observe interactions between motorists and law enforcement conditional on the interaction being recorded (Knox et al. (2019)). As researchers, we cannot know how many motorists a law enforcement officer chooses not to pull over/interact with. Conditional on having the interaction recorded, we also rely on the assumption that the officer records the data as factually as he or she can. Given that a law enforcement officer is biased, he or she is motivated to mis-record the interaction to appear less biased.

Electronic copy available at: https://ssrn.com/abstract=3357063

Traditionally, empirically proving observed disparities in treatment of citizens is driven by racial bias is difficult for a few reasons. First, the true proportion of guilty citizens in the sample is unobservable. Researchers do not know the true number of drivers who actually carry contraband. Second, many unobservable variables, such as driver demeanor, can affect the outcome of trooper-driver interactions, are unavailable to researchers. Without this data, researchers have no way for controlling for all these omitted variables, making it very difficult to understand the true motivation behind racial profiling using the available data. Researchers and economists have attempted to answer this question in recent years, by focusing on the stylized interaction between motorists and law enforcement.

In this paper, I develop a new statistical test of racial bias that exploits the fraudulent behavior by Texas highway troopers in motorist stops from 2010 - 2015. Troopers were caught purposefully misrecording minority motorists' race as white in a stop, a practice which I call masking or cheating. I exploit this masking behavior to develop a new statistical test which links masking behavior to the magnitude of racial bias of individual troopers. The intuition of my test is that troopers mask to cover up biased behavior. Specifically, I find that troopers were masking minority motorists as white in order to make their minority search success rates appear higher than in reality. Masking gives the trooper direct payoff and also a way to cover up racially biased behavior. I build upon past models of Knowles et al. (2001) (which I'll refer to as KPT) and Anwar and Fang (2006) (which I'll refer to as AF) and incorporate this new dimension of masking behavior of troopers. Using this test, I develop a new test and method of measuring racial bias, which yields clues of trooper preferences and motivations.

This paper is related to the literature on motorist stops and whether the decision to search is influenced by racial bias. There are many earlier contributions to the literature that examine the role of motorist race and trooper race in stop interactions, notably KPT and AF along with Antonovics and Knight (2009). None of the past papers have addressed the possibility of the data being purposefully mis-recorded to hide bias. My paper is the first to my knowledge to test for masking behavior in troopers and to explicitly link this behavior to racial bias.

As is well known, racial disparities in aggregate statistics is not evidence of racial bias. For example, if Hispanic motorists are more likely than white motorists to carry contraband, then the aggregate number of searches and stops for Hispanic motorists would be higher even if race was not a factor in the decision. Moreover, troopers attempting to maximize successful searches might racially profile motorists. Such *statistical discrimination*, which is legal in the U.S., race is an indicator for whether the motorist is actually guilty. Thus distinguishing if a motorist search is driven by statistical discrimination or racial bias of the law enforcement officer becomes the fundamental question for policy makers and researchers.

In my model, I use the distortion in the search success rate from cheating to create a measure of bias. I also apply different statistical tests to determine if this masking is intentional. In the context of this

Electronic copy available at: https://ssrn.com/abstract=3357063

research, troopers used cheating to make their white motorist search success rate lower than the actual rate and to make their minority search success rate higher. I use the frequency of cheating and the effect cheating had on the trooper's success rates to determine the trooper's preferences towards motorist rates. My test can not only detect the existence of bias, but can also measure the magnitude of the bias for the individual trooper.

I test my model on a unique data set of 8 million stops created by combining Texas stop data from the Stanford Open Policing Project (SOPP) with administrative data on Texas highway troopers from the Texas Department of Public Safety for 2010 - 2015. I use multiple data driven approaches that leverage the surname and home address of the driver to uncover masking in millions of recorded interactions between troopers and motorists. I use the level of misrecording of race to develop my measure of masking, or cheating, which I then use to measure bias. My research is the first to use this analysis in the context of highway stops.

Using this data, I am able to make a few contributions to the literature on racial prejudice in law enforcement. First, I find that all troopers, including Hispanic troopers are biased against Hispanic motorists. Further, I find that Hispanic troopers are the most biased overall and black troopers are least biased. Troopers are more biased in their interactions when the motorist is guilty. I strictly define guilty as when a trooper discovers contraband (i.e drugs, illegal arms) conditional on searching a motorist. I also use this biased behavior to uncover trooper preferences towards motorist race. Specifically, I find that Hispanic troopers exhibit same race bias by being biased against Hispanic motorists, white troopers are biased against Hispanic and black motorists, and black troopers are rather unbiased. Comparing my results to past work by Anwar and Fang (2006), Knowles et al. (2001), and Antonovics and Knight (2009), my results are more informative about the magnitude and direction of the bias, using this innovative test. Unlike past papers, I find the existence of like-race bias for Hispanic troopers in motorist stops.

Second, I am able to use my measure of bias to uncover covariates of bias such as the role of institutions and the relative proximity to the Mexican border. I also test how bias affects labor market outcomes for troopers. I find that biased troopers have a higher search success rate compared to unbiased troopers, but this gap falls when searching Hispanic and black motorists. I also find evidence of both institutional bias and 'bad apple' troopers when testing for clustering of bias and biased troopers.

Finally, these results motivate testing for accuracy in the data. The issue of misrecording in law enforcement has been heightened in recent years with instances across the U.S., not only in Texas. The increasing number of cities requiring the use of body camera footage during police stops is in response to the concern of the accuracy in police reporting. This paper shows that without correcting for the misrecording, the amount of bias researchers can detect is severely under-measured. Further, by uncovering

Electronic copy available at: https://ssrn.com/abstract=3357063

this bias, researchers can also use this to create a better test of racial bias.

The rest of the paper is organized as follows. In Section 2, I outline the background of my research. Section 3 outlines my statistical model. Section 4 shows my empirical results and other testable implications of my model. I run robustness checks in Section 5 or conclude.

2 Background

Literature Review

One approach that much of the past literature has used to distinguish between racial prejudice and statistical discrimination is to use Becker's (1957) outcome test. In the context of motorist searches, the intuition is simple: if troopers are profiling minority motorists due to racial bias, then they will continue searching minority motorists even if the likelihood of the motorist carrying contraband is smaller than the likelihood for whites. In other words, if racial prejudice is the reason for racial profiling, than the success rate of the marginal minority motorists will be lower than the success rate for the marginal white motorist. On the other hand, if the racial profiling is the result of statistical discrimination, than the search success rates should be same. The intuition being that the trooper will expend his resources searching the race of motorist that is more likely to be guilty. While the approach is simple, the application of this test is difficult. The researcher can never observe the marginal motorists since this requires knowing all the variables that could influence a trooper's decision to search.

Past empirical studies have acknowledged this issue when testing for racial bias (Anwar and Fang (2006)). Even with rich data, researchers cannot prove conclusively the direction of racial bias, even if they can test for the existence (Knowles et al. (2001), Roland G. Fryer (ming), Anwar and Fang (2006)). Without being able to observe the marginal motorist, researchers cannot definitively say if troopers are biased against a certain race of motorist. Another approach is to assume the trooper's preference and then test to see if this holds empirically (Antonovics and Knight (2009)). The main issue with this is that this can only test for relative bias and not absolute bias.

Other researchers have used alternative identification strategies to overcome the selection bias of troopers in the choice of searching motorists. One example is the 'veil of darkness,' which uses the diminished ability of trooper's to observe the motorist race after sunset. While this reduces the prevalence of selection bias, Kalinowski et al. (2017) and Horrace and Rohlin (2016) still find evidence of endogeneity. West (2018) use the plausibly exogeneous assignment of police officers to traffic accidents to identify a causal relationship between the actions of police officers by driver race. The major drawback to this identification strategy is that the results are context specific and may not apply to a wider range of

Electronic copy available at: https://ssrn.com/abstract=3357063

motorist and officer interactions.

My research is also related to empirical research in cheating behavior. Most of these past papers are in the context of cheating behavior in education, whether in teachers and administrators or in students (Jacob and Levitt (2003)). Jacob and Levitt (2003) found that teacher's cheating behavior was highly responsive to incentives. Schools with high-powered incentives induced cheating behavior. This type of motivation and incentives is key to understanding why troopers may choose to mask motorists in stops.

One past paper that studies law enforcement officer using trooper discretionary power in stops to expose racial bias is in the context of providing citations to speeding drivers. Goncalves and Mello (2017) finds that officers are more likely to be lenient when ticketing speed violations with white drivers compared to non-white drivers, which they argue is proof of biased behavior. This is vastly different from my paper in that I study stops conditional on searches and this cheating behavior is less consequential compared to misrecording races. Given that mine is possibly the first to document such severity of cheating behavior in troopers, this may provide motivation to reexamine past work in racial bias in motorist stops. With masking, past literature may still be under-detecting the existence of bias.

Masking and Highway Troopers in Texas

Texas Highway Patrol is a division of the Texas Department of Public Safety. The patrol's primary duties are to enforce state traffic laws and commercial vehicle regulation on highways of Texas. They currently employ over 2800 troopers in Texas divided across 6 regions in Texas, with a separate region for their headquarters in Austin. The department is responsible for licensing of drivers, vehicle inspections, and handgun licensing.

To become a trooper, a person must complete Recruit School or transfer from prior law enforcement service. New hires spend some at least one year as probationary troopers before being permanent assignments. After the one year probationary period, troopers take their final exam and are promoted to trooper.

Troopers are promoted to different levels trooper classes every four years which include salary increases. These trooper classes include Trooper I - V and senior trooper. To qualify for these promotions, a trooper must have had no disciplinary action in the last four years, not been demoted, and pass the physical readiness test. With each salary promotion, troopers can be moved to different stations across the state to fill availability. Troopers are allowed to have more say in the choice of where they are stationed after significant changes in DPS in 2012. After four years of services, troopers can be promoted to corporal. Promotion to corporal is based on availability, similar requirements for promotion as troopers, along with 'demonstrating ability, devotion to duty, and a strong desire to further goals of

Electronic copy available at: https://ssrn.com/abstract=3357063

the department.' To be selected as corporal requires an interview with one supervisor, a corporal, and one minority board member. Unlike other state police agencies, Texas state legislature sets the salary of public employees, rather than individual agencies.

Due to Texas' proximity to the Mexican border, Texas Highway Patrol heavily participates in increased law enforcement along the Texas-Mexico Border. Since 2014, DPS has sent troopers from across Texas to the border to serve for approximately one week. The main goal of the operation was to reduce drug trafficking and undocumented immigration across the Texas-Mexico border.

In a motorist stop, troopers are allowed to investigate the passenger and the driver. While drivers are not required to answer questions, they are required to provide their driver's license and if arrested, they must also provide their name, residence address, and date of birth. Law enforcement officers may ask for consent to search the vehicle or person, which the driver can grant or deny. "... however, if an officer has probable cause to believe that your vehicle contains evidence of crime, it can be searched without your consent (DPS (DPS))." To search a vehicle without the driver's consent, the trooper must either have: probable cause, arrested the driver *prior* to searching the vehicle, reasonably believes the motorist has weapons, or has a warrant. If the officer believes that the driver or passenger has a weapon, he or she may pat down the person and search the vehicle and the surrounding immediate area. Motorists cannot physically resist a search but can notify the officer that he or she does not consent.

Drivers can report troopers if they feel that troopers behaved inappropriately during a stop and troopers can face repercussions if the claim is substantiated. Troopers badge numbers and names are normally provided and drivers can submit complaints to the department. Upon receipt of a complaint, the department assigns the complaint either he Personnel Complaint Investigations or Division Referrals to investigate the complaint. The investigation can have one of four outcomes: unfounded, exonerated, not sustained, or sustained. A sustained complaint can result one or more of the following: formal written reprimand, disciplinary probation, time off without pay, reduction of salary rate, demotion, and or discharge. A formal complaint "alleges one or more of either an infraction of Department rules, regulations, or polices, or an illegal act (TxDPS (2018))." Racial profiling is considered an illegal act under Article 2.132 in the Code of Criminal Procedures and can be a legitimate reason to file a complaint against the trooper.

In 2015, a KXAN investigation found that DPS troopers were " inaccurately recording the race of large numbers of minority drivers, mostly Hispanic, as white" (Collister (2015b)). After this was uncovered, the the House Committee on County Affairs held a hearing where DPS blamed the error on a computer glitch. As a result of the hearing, the percent of White motorists being stopped fell from 18% to 4% (Collister (2015a)).

An important result of the KXAN investigation was that masking was also found in other law en-

Electronic copy available at: https://ssrn.com/abstract=3357063

forcement departments in Texas, namely Houston and Austin police departments. This is important as many previous studies do not account for possible cheating behavior in police or trooper forces. This raises the question if whether past reports and research of racial bias are possible under-measuring and under-detecting the existence of bias. Texan troopers now ask drivers for their race rather than using information off of the driver license or their own best judgment (Oyeniyi (2015)).

Masking is easy in motorist stops compared to other points of the criminal justice system. First, the trooper is not required to ask the driver for his or her race. Instead, the trooper is supposed to infer the race based on observable characteristics of the driver. Second, due to the high frequency of stops, troopers or police officers who participate in masking are not checked for accuracy and are less likely to be caught. Usually, only the driver focuses on the content of the ticket. Third, unless the trooper searches the driver and arrests the driver, no other law enforcement officer will see the recorded race.

3 The Model

In this section, I present a straightforward model on trooper search behavior that yield novel tests of racial bias. Suppose we have troopers and motorists; motorists are of race $m \in \{M, W\}$ and each individual trooper of race-$t \in \{M, W\}$. Suppose that among motorists of race m, a fraction π^m choose to carry contraband. This information is available to the trooper along with other pertinent characteristics that are collapsed to a single index $\theta \in (0, 1)$.[*] If a driver of race m is actually carrying contraband, then $\theta \sim f_g^m(.)$; if the driver isn't carrying contraband, then $\theta \sim f_n^m(.)$. I assume that the two densities are continuous and satisfy the strict monotone likelihood ratio property. Intuitively, this property implies that a higher index of θ implies a higher probability of driver guilt.

3.1 Search and Masking

Each trooper of race-t can choose to search a motorist after observing the motorist's vector of characteristics, (m, θ). I assume that a trooper wants to maximize the number of successful searches (searches where illicit contraband is found). When a race-t trooper searches motorist of race-m, she incurs a cost of $c_{m,t}$. If the driver is guilty, the trooper receives a benefit, normalized to one such that the cost of the search, $c_{m,t} \in (0, 1)$.

Let G denote the event that a motorist is guilty of carrying contraband. When a trooper pulls over motorist, she observes m and θ. The ex-ante probability the motorist is guilty conditional on the observed

[*]Some examples of these characteristics are age, height, address, gender, the interior of the vehicle, the smell of the driver, whether the driver is under the influence, whether the license plate is in-state, the time and place of the stop, whether the vehicle is rented, and the attitude of the driver.

Electronic copy available at: https://ssrn.com/abstract=3357063

m and θ is:

$$\Pr\left(G=1|m,\theta\right)=\frac{\pi_m f_g^m\left(\theta\right)}{\pi_m f_g^m\left(\theta\right)+\left(1-\pi_m\right)f_n^m\left(\theta\right)} \tag{1}$$

From the Monotonic Likelihood Ratio Property, $P(G|m,\theta)$ is monotonically increasing in θ.

The trooper then decides to search the motorist of race-m and signal θ based on the expected payoff of searching such that:

$$max\{P(G|m,\theta)-c_{m,t};0\} \tag{2}$$

The first term is the expected benefit of searching that motorist and the second term is the payoff for not searching. Therefore, the optimal decision for a trooper of race-t to search a motorist of race-m with observed signal θ if and only if:

$$\Pr\left(G=1|m,\theta\right)\geq c_{m,t} \tag{3}$$

The trooper has a search threshold θ^* where (3) holds with equality.

Similar to Knowles et al. (2001) and Anwar and Fang (2006)I define two types of racial prejudice. A trooper exhibits racial prejudice if she has a taste or preference for searching motorists of a certain race, which is modeled using the search cost, $c_{m,t}$.

Definition 1. A trooper of race-t exhibits naive racial bias against motorist of race m if $c_{m,t}<c_{m\prime,t}$.

Next, I define statistical discrimination for troopers if they have no taste for racial bias, but still use a different search criteria for motorists of different races.

Definition 2. An unbiased trooper with $c_{m,t}=c_{m\prime,t}$ exhibits statistical discrimination against race m motorist if $\theta^*(m,i)\neq\theta^*(m\prime,i)$ for $m\neq m\prime$

After observing the search outcome, G, the trooper decides whether to mask the motorist of race M as a motorist of race W. Masking incurs a cost of $\mu(\theta)$, which is a function of all observable characteristics of the motorist, aside from race. If the motorist is guilty, $\mu(\theta)>0$ and is increasing in θ.; if the motorist is not guilty, $\mu(\theta)<0$ and increasing in θ. The trooper will mask if and only if:

$$c_{M,t}\geq c_{W,t}+\mu_{MW,t}(\theta) \tag{4}$$

which only holds if the motorist is not guilty.

By masking the M motorist as W, the trooper incurs the cost, $c_{W,t}$ plus the masking cost $\mu_{MW,t}$. The trooper has a masking threshold, $\theta^{\mu}_{M,t}$ s.t (4) holds with equality. And,

$$\theta^*_{M,t}<\theta^{\mu}_{M,t}$$

Electronic copy available at: https://ssrn.com/abstract=3357063

Definition 3. A race-t trooper exhibits sophisticated racial bias against motorist of race m if $\mu_{m,m\prime,i} < 0$ for $m \neq m\prime$

3.2 Equilibrium

For the rest of the section, I assume that only race M motorists are masked as W. The *equilibrium search rate* of trooper t against race M motorist, $\gamma_{M,t}$, is given by:

$$\gamma_{m,t} = \pi_m[1 - F_g^m(\theta^*)] + (1 - \pi_m)(1 - F_n^m(\theta^*)] \tag{5}$$

For a race t trooper, the *equilibrium search success rate* is :

$$S_{m,t} = \frac{\pi_m[1 - F_g^m(\theta^*)]}{\pi_m[1 - F_g^m(\theta^*)] + (1 - \pi_m)[1 - F_n^m(\theta^*)]} \tag{6}$$

With masking, the true search rates and the true search success rates are unobservable to the trooper. I will assume for the rest of the section that troopers only mask race M motorists as race W. Therefore the *observed search rate* for race M motorist is:

$$\gamma^O_{M,t} = \pi_M[1 - F_g^M(\theta^\mu)] + (1 - \pi_M)[1 - F_n^M(\theta^\mu)] \tag{7}$$

For race t trooper t, the *observed* search rate for race W motorist is:

$$\begin{aligned}\gamma^O_{W,t} = \pi_W[1 - F_g^W(\theta^*)] + (1 - \pi_W)(1 - F_n^W(\theta^*)] \\ +(1 - \pi_M)[F_n^M(\theta^*) - F_n^M(\theta^\mu)]\end{aligned} \tag{8}$$

For race t trooper t, the *observed* search success rate for race M motorist is:

$$S^O_{M,t} = \frac{\pi_M[1 - F_g^M(\theta^\mu)]}{\pi_M[1 - F_g^M(\theta^\mu)] + (1 - \pi_M)[1 - F_n^M(\theta^\mu)]} \tag{9}$$

For race t trooper t, the *observed* search success rate for race W motorist is:

$$S^O_{W,t} = \frac{\pi_W[1 - F_g^W(\theta^*)]}{\pi_W[1 - F_g^W(\theta^*)] + (1 - \pi_W)(1 - F_n^W(\theta^*)] + (1 - \pi_M)[F_n^M(\theta^*) - F_n^M(\theta^\mu)]} \tag{10}$$

For race t trooper t, the masking rate for race M motorist is:

$$\phi_{M,t} = \frac{(1 - \pi_M)[F_n^M(\theta^\mu) - F_n^M(\theta^*_M)]}{\pi_m[1 - F_g^m(\theta^*)] + (1 - \pi_m)(1 - F_n^m(\theta^*)]} \tag{11}$$

Electronic copy available at: https://ssrn.com/abstract=3357063

The second definition of racial bias is based on the masking rate of trooper t. Given the trooper is not Type 1 bias, she will not mask at all. From Definition 1, if the trooper is unbiased, then $c_{m,t} = c_{m\prime,t}$. Therefore, there is no $\mu_{mm\prime,t}$ s.t eq(4) holds.[†]

Suppose a trooper, t is biased against race-M motorists compared to race-W motorist. Then $c_{M,i} < c_{W,i}$ and there exists a $\theta^{\mu}_{MW,t}$ s.t eq (4) holds. Then we can test for bias by simply testing if the trooper masks at all, or if $\eta_{M,i} = 0$ and $\eta_{W,i} = 0$.

3.3 Testable Implications

Now I derive some simple tests of the model that I will also test empirically. First, if the troopers are not biased such that $c_M = c_W$, then there is no masking threshold, $\theta^{\mu}_{MW,t}$ such that $c_{M,i} > c_{W,i} + \mu_{W,i}(\theta)$.

If the trooper is racially biased against race-m motorist, then the assumptions of the MLRP provide an intuitive test of racial bias. From the MLRP, since the $\frac{f^m_g}{f^m_i}$ is strictly increasing in θ and $\frac{f^m_g}{f^m_n} \to 0$ as $\theta \to 1$. The MLRP also implies that the cumulative distribution function F^m_g first order stochastically dominates F^m_n

$$\Rightarrow F^m_g(\cdot) > F^m_n(\cdot)$$

$$\Rightarrow 1 - F^m_g(\cdot) < 1 - F^m_n(\cdot)$$

Proposition 1: Detecting bias: If a trooper is racially biased against race-M motorist, then $\phi_{M,i} > 0$ s.t:

$$\phi_{M,t} > 0 \tag{12}$$

Next, it is important to be able to test for the direction of the bias.

Proposition 2: Direction of bias: If a trooper t is racially prejudiced against race M motorist, then he/she will mask race-M motorist as W.

This combined with Proposition 1 also implies that the observed search success rate for race-M motorists will be higher than the actual search success rate since the trooper will only mask unsuccessful searches. Suppose a trooper t masks minority (M) motorists as white (W),

$$S^O_{M,i} > S_{M,i}$$

and

$$S_{W,i} > S^O_{W,i}$$

$$\Rightarrow S_{M,i} - S^O_{M,i} < 0; S_{W,i} - S^O_{W,i} > 0$$

[†]Note, all troopers with Type 2 racial bias are also Type 1 biased, but not all Type 1 bias troopers are Type 2.

Electronic copy available at: https://ssrn.com/abstract=3357063

To measure the magnitude of bias, I use the masking rate. Suppose trooper i and trooper j are biased against race M motorist, but trooper i is more biased such that $c_{M,i} < c_{M,j}, c_{W,i} = c_{W,j}$, and $c_{M,t} < c_{W,t}$ for $t \in \{i, j\}$. Since both troopers face the same population of race-M motorist and race-W motorist, then this implies that $\theta^{\mu}_{M,i} > \theta^{\mu}_{M,j}$ and $\theta^{*}_{M,i} < \theta^{\mu}_{M,i}$. From formula (6), formula (9), formula(10), and formula (12), this implies that:

$$\phi_{M,i} > \phi_{M,j}$$

and

$$\Rightarrow S_{M,i} < S_{M,j} \text{ and } S^{O}_{M,i} > S^{O}_{M,j}$$
$$\Rightarrow S_{M,i} - S^{O}_{M,i} < S_{M,j} - S^{O}_{M,j}$$

Proposition 3: Magnitude of bias: If a trooper t is racially prejudiced against race M motorist, then the magnitude of bias is simply:

$$\phi_{M,t} = Masking\ Rate \tag{13}$$

Suppose the trooper of race t is naively biased against race M motorist such that $c_M < c_W$ while $\eta_{M,t} = 0$. Then from Anwar and Fang (2006):

Proposition 4: If neither race M nor race W troopers exhibit Type 2 racial prejudice, then the ranking of $\gamma_{m,M}$ and $\gamma_{m,w}$ nor the ranking of $S_{m,M}$ and $S_{m,W}$ depends on $m \in \{M, W\}$.

3.4 Data

3.4.1 Stop Data

The Stanford Open Policing Project (SOPP) has collected over 130 million records from 31 state police agencies (Pierson et al. (2017)). The goal of the project is to analyze detailing interactions between police and the public. This information is freely available on the website.

I only use the Texas portion of the SOPP data. While SOPP provides the data from 2006, Texas troopers were not required to record the driver's last name until 2010, so I cannot test for masking behavior prior to 2010. This data contains detailed information on the stop such as latitude and longitude of the stop, time and date of the stop, the reason for the stop, whether a search was conducted and why, if contraband was found, whether an arrest was made, first initial and last name of the trooper recording the stop, and the badge number of that trooper. The data set also has limited information on the type of contraband found: currency, weapon, and other. The unique feature of this data set is it also contains detailed information on the motorist such as: driver's first and last name, address of the driver, recorded

Electronic copy available at: https://ssrn.com/abstract=3357063

race of the driver, make and model of the car, the owner of the car. This becomes important when I do the race correction. I drop Native American, Asian, and Middle Eastern motorists, which is about 1 million of the observations. I also drop stops where the trooper did not record the race of the driver. For reasons I explain in the race correction analysis section, I also only keep Texas, male drivers. Overall, the subset of the data I use contains about 8 million total stops with 3,509 unique troopers.

In Texas, troopers can legally search a vehicle for many reasons aside from probable cause or driver consent. Some of these situations, such as search incidence to arrest, after the car is impounded, or with a warrant, do not fit the framework of the model. One of the assumptions in my model is that motorists are only guilty through finding contraband. If the motorist is arrested prior to searching the vehicle, then that will bias my results. I restrict my definition of search success to only include searches due to probable cause or driver consent.

I also supplement the SOPP data with 2016 - 2017 highway stop data from the Texas Department of Public Safety. This data has identical information to the SOPP data, but does not have the driver's full name or addresses. The new data set contains additional information such as whether the driver was a fugitive, the sergeant area of the stop, the alleged speed, the judge assigned to the case, and the court date and location. I also drop the female driver's from this data set. Since the stops occurred after the masking was revealed in November 2015, I take the driver's races as given.

3.4.2 Race Correction

I use two main methods supported by past literature on using observable characteristics to determine race. These methods are predominantly used in social science and health research to infer patient race (Fiscella and Fremont (2006)). The first method is to use surname analysis, which works well for Hispanic and Asian surnames. I match the driver surnames in my data to the U.S. Census Surnames data set. If the probability of the last name is Hispanic is greater than a certain threshold (75%), I impute the 'corrected' race as Hispanic.[‡] For example, Figure 6 shows an actual ticket from a stop. The driver, Antonio Tovar Mendez, is pulled over for speeding by Officer Salinas and is recorded as a white Male driver. Using the surnames analysis, the probability Antonio is Hispanic, conditional on his last name, Mendez, is 92%. I then correct his race to Hispanic. The advantage of this method is that the correction is fairly quick and simple. But, the main drawback is that this method is only suitable for Asian and Hispanic names and is less effective with females. Thus, I only keep male drivers in my sample.

The second method I employ is geocoding analysis. This method works best for black drivers "because at least half of black Americans continue to live in predominantly black neighborhoods (Fiscella and Fremont (2006))." I use the recorded address of the driver to geocode to a specific latitude and longitude

[‡]As a robustness check, I raise the threshold to higher levels

Electronic copy available at: https://ssrn.com/abstract=3357063

using geocoder.us. I then use that latitude and longitude to map the address to a Block FIPS code using the FCC block finder. I merge this data with the 2010 American Community Survey. If the percentage of Black population in the area is greater than a certain threshold (75%), I correct the race as "black." This method also has a few disadvantages. First, if the trooper did not record the address of the driver ($< 7\%$ of the data), I can't geocode it. Second, the address is inputed by the trooper, which is prone to spelling and typing errors. For example, I found 116 different spellings of the city "Houston," which is the largest city in Texas. Third, this method is also very computationally expensive so I restrict this analysis to only drivers who live in Texas, which is almost 90% of the stop data. [§]

The effect of the surname and geocoding analysis on the aggregate statistics for stops is shown in Table 2 and in Figures 3 - 5. Using these methods, I correct over 1,270,734 of stops as Hispanic, nearly 11,669 number of stops as black, and 1,121,133 as white. This has rather large effects in the aggregate stop statistics for Hispanic and white motorists as nearly 1 million white stops are actually from Hispanic motorists. This also has an effect on the search rates as in the observed statistics, Hispanics are searched only 17.4% of the time while after correction, the search rate more than doubles to 42.1%. The difference between the observed and corrected rates for non-white motorist and white motorists do not add up as some troopers simply reported the motorist race as Unknown rather than putting down a race. Of the 161,270 Unknown race, 160,909 I correct as Hispanic and 316 I correct as Black.

In Figures 4 - 3, I show the time trend of motorist stops with the observed driver's race and the corrected driver's race. From January 2010 to May 2015, I observe an increase in the number of Hispanic stops and a decrease in white stops, implying a change in masking behavior of troopers or a change in the driver population in Texas. I also find that after June 2015, the observed Hispanic stop rate fell significantly from 30% to 10%, with spikes in the white motorist stop rate and black stop rate. This implies that troopers were masking Hispanic motorists as white and black. I found no news articles to explain this change in masking behavior so I omit the data prior to May 2015. I only correct the races of motorists originally recorded as white or unknown. This allows me to only correct the races of the motorist once. Otherwise, a motorist with the surname Gomez living in a predominantly census block could be corrected as black or Hispanic depending on if I ran the surname analysis first or the geocoding analysis. I will go into further detail my methodology for correcting the race of the motorist in the subsequent section

3.4.3 Trooper Employment Data

The employment data is from the Texas Department of Public Safety. Unfortunately, DPS only has this information for employees during 2013 - 2015. If a trooper left DPS prior to 2013, I do not have his or her employment information. For troopers in the data, I have the year the trooper was hired, if he or she

[§]I also raise this threshold later as a robustness check

Electronic copy available at: https://ssrn.com/abstract=3357063

left the position and why, the salary for each year, which work city he or she was stationed at, the work position for each year, ethnicity of the trooper, the full name of the trooper, and the badge number. The key limitation for this data is that I do not have employment data prior to 2013. I have approximately 2,788 unique troopers of which I can match 2,578 to the stop data.

I merge these two data sets together using the badge number of the trooper. I can match all but 10% of the stop data to the trooper so I only have 11,819,236 observations. If I match the trooper, then I use the information in the earliest matched year to match to prior years, which is appropriate since certain traits of the trooper, such as sex, name, and race, do not change across time. I also drop troopers who have less than 4 searches from 2010-2015. My final number of observations is 5,882,679 after dropping observations after May 2015.

3.5 Descriptive Statistics

I present summary statistics of motorist characteristics in Table 2 using the corrected data. On average, most motorists stopped are white, but this pattern doesn't carry over to searches. Instead I find that conditional on being stopped, Hispanics motorists are searched the most at 42% followed by white motorists at nearly 40%. Black motorists also show a higher search rate compared to stop rate with a difference of 8%. I also find that certain stop characteristics, such as having a Midnight stop, an older car and a luxury brand card are also more likely to be searched compared to the stop rate.

Table 3 shows summary statistics of troopers. Of the 2,591 troopers I was able to match to the data, approximately 61% are white, 30% are Hispanic, and almost 9% are black. I drop Asian troopers, which is less than 1% of the data. The force is predominantly male at 94%. By trooper race, I find that white troopers are most likely to search at 1.6% of the time, followed by Hispanic motorists at 1.3%. For masked stops, I find that Hispanic motorists are more likely to mask at 31% while black and white troopers only mask at over 14%. From Figures 1 and 2, Hispanic troopers are more likely to be placed in areas with high Hispanic stops, which are close to the border. I exclude border counties in column 4 and I find that Hispanic troopers make up only 25% of the masked stops in non-border counties. I find similar drops for black and white troopers, which fall by 2% down to 15%.

In the bottom part of the table, I break down the stop and search statistics by trooper position. Ranked officers make up only 20% of the highway patrol. The rank of the officer does not have any significant effect on the search rate, but sergeants are significantly more likely to mask in stop when compared to those ranked below. Using the rank of Captain as an example, the interpretation of the probabilities is "If the trooper is a captain, then the search rate is 5%." I observe similar probabilities for searching regardless of trooper rank between 1.4-2% except for Captains at less than 1%. For masked

Electronic copy available at: https://ssrn.com/abstract=3357063

stops, I find that higher rank is correlated with more masking with the exception of Captains. Captains mask 10.6% of the stops whereas troopers mask 22.3% of the time.

4 Empirical Results

Past tests for racial prejudice

Before showing the results of my test of bias on the data, it is important to highlight the effect of masking on past statistical tests of bias, notably, Knowles et al. (2001). In Table 4, Column 1 shows the results of Knowles et al. (2001) using the observed, raw data. Under their test, I would find bias against Hispanic and white motorists compared to black motorists. Since Hispanic motorists have a lower success rate of 31.9% compared to white search success rate at 42.0%, I would conclude that Hispanic motorists were the most biased against. After applying the race correction, I find different results from the test. Using KPT's test but on the corrected data, I find that both black and Hispanic motorists are biased against and that the magnitude of bias against Hispanics is actually much larger. Without taking in account the race correction, KPT's test would incorrectly conclude bias against Hispanics and whites, rather than against Hispanics and blacks.

This also shows that masking reduces the appearance of racial bias for troopers by reducing the search success rate of white motorists. While I find no large increases in the search success rate after race correction for black and Hispanic motorist search success rate, the white search success rate has risen significantly by 7% to nearly 50%. This implies that trooper's also differentially mask based on search outcome, which becomes the basis of my racial bias test.

My test for racial prejudice

Thus far, I have shown strong evidence that masking not only distorts the observed stop and search rate of motorists, but it also leads to the wrong conclusions of tests for racial bias. Trooper's were systematically misrecording races of motorists, especially Hispanic motorists to appear less likely to stop and search Hispanic motorists. Using my model, I show this behavior is linked to bias by comparing the rate of masking conditional on the search outcome and by comparing the search success rate using the observed data and the corrected data. In Table 5, I show the results of my main test of bias, which compares the probability of masking conditional on search outcome. The key identification of my test is that biased troopers should differentially mask based on search outcome; specifically, the trooper should mask only when the search ends in failure. Since my race correction method cannot perfectly detect motorists race and is prone to its own errors, I rely on the differential masking behavior across search

Electronic copy available at: https://ssrn.com/abstract=3357063

outcome to measure bias. If the trooper is biased, then I should find higher rates of masking when the search ends in failure compared to when the search ends in success.

Table 5 shows the masking rate conditional on search outcome in columns 1 and 2. If troopers are indeed masking in a biased manner, then the $Pr(Mask|Failure) > Pr(Mask|Success)$ according to my model. Biased troopers should only mask searches that end in failure with the intuition that masking successful searches raises the probability of be being caught masking. Indeed, I find that troopers are 0.5% more likely to mask black motorists as white when the search ends in failure compared to when the search ends in success, which is a small, but significant difference. For Hispanic motorists, I find that failed searches are 3.3% more likely to be masked compared to successful searches.

To measure the magnitude of bias, I use Eq. (13):

$$\phi(R, t|G = 0) - \phi(R, t|G = 1)$$

where R stand for the driver race, and t stands for trooper race. The more biased the trooper is, the more he will mask motorist of race-R when the search ends in failure, $G = 0$, compared to when the search ends in success, $G = 1$. I rely on the difference because my method of race correction also corrects successful searches so the differential rate of masking based on search outcome will identify bias.

Using Eq. (13), I create a measure of bias for each individual trooper. In order to ensure that troopers who only search a few times during the time period do not bias the estimates, I only use troopers who have conducted at least four searches from 2010 to 2015. Figure 8 shows the distribution of my estimated measure of trooper bias against Hispanics. Officers on the positive side of the histogram exhibit racial bias against Hispanic motorists and officers to the left of the histogram are unbiased. The interpretation of the values on the x-axis is the difference in masking behavior across search outcomes. For example, at 0.5, the officer is 50% more likely to mask when the search ends in failure compared to when the search ends in success.

One notable characteristic of the distribution of bias is the heterogeneity in the measures of bias for troopers. While most troopers are concentrated at no bias, I found a large mass with positive levels of bias, with a mean and a median of approximately 0.2. Officers at the 90th percentile of the distribution are 66 percent more likely to mask with failed searches compared to successful searches while at the bottom 10th percentile are 25% more likely to mask successful searches than failures.

Another important fact of the distribution is the negative side. Troopers here are more likely to mask when the search ends in success compared to failure. This can occur for the following reasons. First, some troopers may only mask when the search ends in success for reasons that may or may not be related to bias. If that is the case, the very few troopers engage in this behavior, as evidenced by the

Electronic copy available at: https://ssrn.com/abstract=3357063

small mass on the left hand side of the distribution. Second, since I rely on the difference in masking behavior, the measure of bias is very sensitive to the number of searches by search outcome and masking. On average, troopers on the left side of the distribution have only 29 number of searches, while troopers to the right of the distribution have 38 searches. Lastly, the race correction method itself is prone to its own errors. Given its imperfections, it would be suspicious if I had no troopers behave opposite of the model predictions.

Robustness Checks

To ensure that my measure of bias is truly capturing bias, I compare my measure of bias to KPT's test of bias. If troopers are biased using my measure of bias, then according to KPT, they will oversearch Hispanic motorists compared to white motorists. This will lead to troopers biased using my measure having a lower search success rate for Hispanics compared to white motorists. I test this using:

$$\begin{aligned} Y_{ijt} = \alpha + \beta_1 HispBias_j + \beta_2 I(MotoristRace_i) + \\ \beta_3 HispBias_j \times I(MotoristRace_i) + \gamma_t + \phi_c + \gamma_t \times \phi_c + \epsilon_{ijt} \end{aligned} \tag{14}$$

where Y_{ijt} is the outcome of a stop, either probability of search or the probability of search success, of motorist i with trooper j at time t. $HispBias_j$ is the standardized measure of Hispanic bias measured for trooper j. $I(MotoristRace_i)$ is an indicator for motorist race of the stopped driver, i, with white motorists being the excluded category. I also included year fixed effects with γ_t and county fixed effects with ϕ_c and also control for county specific time trends. The two outcomes I am interested are search success and search. The coefficient of interest is β_3, which can be interpreted as the differential treatment of Hispanics motorists compared to white motorists for biased troopers. If my measure of bias is in line with KPT's measure of bias, then troopers biased against Hispanic motorists will have a lower search success rate for Hispanics when compared to white motorists, thus β_3 will be negative.

My results in Table 17 indeed show agreement with KPT's measure of bias. Namely, one standard deviation increase in my measure of bias reduces the probability of search success for Hispanic motorists by 1.3%, with both estimates being significant. I also find that Hispanic bias reduces search success rates overall for biased troopers where one standard deviation increase in bias reduces search success rates by 13.6%. I also find that troopers biased against Hispanics also tend to search Hispanics 0.2% more than white motorists.

I also ensure that the relationship between my measure of bias and KPT's measure of bias is not dependent on my census surname cutoff, I vary the threshold I use in the surname analysis at 50%, 75%

Electronic copy available at: https://ssrn.com/abstract=3357063

(the measure I use throughout my analysis), 85%, and 95%. My results in Figure 10 show that even with tighter thresholds, the troopers who engage in biased masking behavior also have a greater search success rate for white motorists than with Hispanic motorists.

As another robustness check, I use other observable characteristics from the stop to show that my measure of bias is actually measuring trooper bias. One possible concern is that troopers who stop motorists at night are less able to accurately identify motorist race and find contraband, which is why I find higher rates of mis-recorded races for failed searches. I report the results of that test and find no correlation between stops made between 10 PM and 5 am have no correlation with bias for black motorists and I find a negative, small, and significant coefficient for Hispanic motorists.

I test another prediction of my model in Table 7. In my model, troopers base their search decision on observable characteristics of the stop, θ. My model finds that troopers mask in a search when they observe a less 'guilty' signal from the stop. I can use search consent as a proxy for the signal of guilt the officer perceives. If my model is correct, then masked stops are more likely to occur when motorist consents compared to when the trooper has probable cause to search a vehicle. With probable cause, the trooper needs some facts or evidence (example: sees or smells contraband) to legally search a vehicle. I find that for black motorists, I find a positive and significant coefficient, implying my model is correct. But for Hispanic motorists, the coefficient is positive, but small and insignificant. While this could mean the predictions from the model are wrong, it could also be due to troopers misrecording motorist consent. I will specifically test for this in later sections.

Next, I use the publication of the news article by KXAN revealing the masking as a natural experiment. The article was published in November 11th, 2015, a hearing was conducted by November 18th, and by November 23rd DPS changed its policies to require troopers to ask drivers for their race. I can test the affect on stop behavior of troopers after the changes are implemented. Since I observe changes in masking behavior from June 2015 to November 2015, I will only use data preceding May 2015 as my pre-data. For my post data, I am using the publicly available, which has the recorded driver's races. If my measure of bias is correct, troopers who were using masking to hide their bias should have the greatest changes in stop behavior with Hispanic motorists.

To do this, I run a linear probability model where the outcome, H_{jt}, probability of being stopped, probability of being searched conditional on being stopped, and the probability of search success on the measure of racial bias for the trooper is interacted with year dummies:

$$H_{j,t} = \alpha + \beta_0 Hisp\ Bias_j + \sum_{t=2010}^{2017} \beta_t Hisp\ Bias_j \times I(Year = t) \\ + X_{c,t}\gamma + \epsilon_{j,t} \tag{15}$$

Electronic copy available at: https://ssrn.com/abstract=3357063

The primary coefficient of interest is β_t for $t > 2015$, which is the interaction between officer level bias pre-2015 and the years after the changes were implemented. $\beta_{2016}, \beta_{2017}$ will reflect how much more the stop rate, search rate, search success rate of Hispanic and white motorists change with bias. If my measure is capturing bias, then officers with higher levels of bias should change more than officers with low level of bias.

I first show the results for white motorists in Table 9 since masking shifts Hispanic motorists into the search statistics for white motorists. Specifically, biased troopers masked their Hispanic search failures as white, thus reducing the white search success rate. If my measure of officer-level bias measures this behavior, then the search success rate should rise disproportionately for biased officers. From column 1, I find no change in the stop rate after the publication of the article, but I do find a decrease the search rate of 0.2% in 2016 and 0.4% in 2017. Focusing once more on the success rate in column 3, I find the search success rate also rises differentially with the trooper's bias. Specifically, a standard deviation away from the mean level of bias leads to a 6% increase in the success rate in 2016 and a 5.6% increase in 2017.

For Hispanic motorists, without masking, the search success rate for Hispanics should fall. But, since biased officers can no longer use masking to hide their bias, I also test to see if and how they change their behavior in response to the 2015 publication. From Table 8, I find no significant change in the stop rate of Hispanic motorists for officers with higher levels of bias after 2015. But, I find that one standard deviation increase in the officer's level of bias is correlated to a 0.5% fall in the Hispanic search rate in 2017. I find a small and insignificant change in 2016. The search success rate, shown in column 3, is more informative on actual changes in troopers for which I find a drastic difference in behavior for biased troopers after 2015. While the search rate for Hispanic motorists does not change in 2016, the success rate rises by 5% with one standard deviation away from the mean level of bias and by 8% in 2017. These are 13.5% and 21.6% increases in the search success rate, respectively. Without masking, biased officers reduced searching and improved their search ability for Hispanic motorists.

4.1 Bias and Institutions

Another important question is whether troopers are biased because of different social norms within counties or because they are 'bad apples.' This is especially important in the context of Texas since troopers by the border have different incentives (for example: reducing drug trafficking, reducing undocumented immigration, ensuring border protection) to search motorists compared to trooper stationed away from the border. In order to test this, I create two new measures to test for these different types of bias. If

Electronic copy available at: https://ssrn.com/abstract=3357063

different social norms within counties are driving the bias, then most troopers stationed in that county should be biased, thus the average level within the bias will be more shifted to the right. I call this type of bias institutional bias. If bias observed within a county is driven by poor discipline, then I should observe high dispersion of bias within the county. I call this type of bias, 'bad apples.' To illustrate this point, Figure 12 shows two example counties, one with institutional bias and the other with 'bad apple' troopers. For the counties with bad apple troopers, the average level of bias is low, but the variance of officer level bias within the county is wide. For institutional bias, these counties have no outliers but the whole distribution is shifted to the right to have a positive, average level of bias.

Using this intuition, I measure institutional bias as the mean of county's level of bias weighted by the number of searches of each trooper patrolling in the county. To measure 'bad apple' bias, I use the coefficient of variation within a county, which is the standard deviation of bias within the county divided by the mean using the same weights. I test if these differences in social norms across the two types of counties is driven by certain county characteristics, which are measured using data from the American Community Survey. Specifically, I regress:

$$\begin{aligned} S_c = \alpha + \beta_1 PercBlack_c + \beta_2 PercHisp_c + \beta_3 Perc\ no\ health\ ins_c + \\ \beta_4 Perc\ HS\ Diploma_c + \beta_5 Median\ HH\ income_c + \beta_6 Perc\ Employed_c + \\ \beta_7 Perc\ Age > 16_c + \beta_8 Total\ Pop_c + \beta_9 I(Border = 1)_c + \epsilon_c \end{aligned} \tag{16}$$

where S_j is the measure of institutional bias or bad apples within a county. All of the explanatory variables are from the 5-year ACS from 2010 - 2015 except $I(Border = 1)$, which is an indicator variable for if the county is defined as county that borders Mexico as defined by the Le Paz Agreement. A positive coefficient for $\beta_1 - \beta_8$ imply higher instances of that trait are associated with institutional bias or bad apples.

Table 10 shows the results for institutional bias in Column 1 and bad apples in Column 2. I find that counties with higher institutional bias have lower percentage employed and smaller population sizes. This implies that counties with low population and low level employment are more likely to have higher level of bias. A one percentage decrease in employment is associated with a 1.32 standard deviation decrease in the mean level of bias within the county.

For counties with high variation in trooper level bias, I find a similar relationship with employment and population. I find that employment has a much larger effect on the dispersion of bias within the county with a one percent decrease in employment correlated to a 4.4 increase in the coefficient of variation.

Another important question of the role of institutions on bias is determining whether troopers change their behavior based on the social norms within the county. For example, a good trooper behaves unbiased

Electronic copy available at: https://ssrn.com/abstract=3357063

in County 1, but is transferred to County 2, where all troopers mask in a biased manner. Thus, this good trooper changes his behavior to match his coworkers. If this is true, then a trooper's own bias should increase if he moves to a more biased county. To test this effect, I use the two counties each trooper conducts the most searches in and estimate his bias in each county. I take first differences across the two counties using the trooper's own measure of bias and the same measure of county level of bias from Table 10:

$$HispBias_{j,1} - HispBias_{j,2} = \alpha + \beta_1(S_{c,1} - S_{c,2}) \tag{17}$$

where $S_{c,t}$ is equal to the mean level of bias within county c or the coefficient of variation of bias of county c.

Given the significance of the border on institutional bias within the county, I focus my analysis onto troopers who move between border and non-border counties and vice versa. Specifically, if a trooper moves to a county with higher bias, does his bias increase also? To test this, I create officer-county measures of bias for each trooper and his top two most patrolled counties. Out of all the troopers, I find 460 who have more than four Hispanic searches in two counties of which 35 have moved from border to non-border and 112 who have moved from non-border to border. Using the same specification as in 17 I include an interaction for whether the trooper moves from Border to non-border counties and for non-border to Border counties.

Table 11 shows these results using a first difference regression using each trooper's top two most patrolled counties. Column 1 reports the results with no controls for border-nonborder movers. I find that a trooper switching his patrol to a county with a one standard deviation increase in average bias increases the trooper's own bias by 0.48 standard deviation. Column 2 and Column 3 focus on the direction of the moving relative to being a border county. I find a negative and significant coefficient for troopers who from to a non-border county and I find no significant effect for officer bias when moving from a border county to a non-border county. But, I find that the direction of moving relative to the border is irrelevant given the same change in mean level of bias in county since the interactions are of similar magnitude. This implies that the change in average level of bias is the important determinant in changing officer's bias rather than the proximity to the border.

Table 12 describes the relationship between the coefficient of variation within the county and the officer level bias. From Column 1, I find that the change in the coefficient of variation has no effect on officer's own level of bias. Similar to the preceding table, I break out the moving officers into groups relative to the border. Here, I find that only officers who move from border to non-border counties with higher coefficient of variation also increase their own bias. Compared to Table 11, this coefficient is much smaller in magnitude. Troopers who move from non-border to border counties have a lower level of bias,

Electronic copy available at: https://ssrn.com/abstract=3357063

but there is no significant effect when interacted with the change in the dispersion of bias within the county.

Bias and Trooper Characteristics

The next analysis is to understand if any trooper characteristics are related to bias. One major contribution of this paper is to be able to generate trooper-level estimates of discrimination. In this section, I will test whether bias varies by trooper demographics. Additionally I will address how discrimination varies with other employment characteristics such as promotions, salary, and officer transfers.

Table 17 shows column 3 of Table 5 broken down by trooper race. To control for the heterogeneity in motorist stops by counties across Texas and to control for the county specific time trends, I use a regression with county by year fixed effects. I define a biased stop as a search that ended in failure where the driver's race was misrecorded and is considered unbiased if it is a failed stop where the race was correctly recorded. I run the regression:

$$BiasedStop_{ijt} = \alpha + \beta_1 Hisp\ Troopers + \beta_2 Black\ Troopers + X_{c,t} + \epsilon_{ijt} \tag{18}$$

where $BiasedStop_{ijt}$ is the dependent variable defined from before. The variables of interest are β_1 and β_2, which are indicator variables for the trooper's race with white being the omitted category. To control for the county specific time trends, I also include county by year fixed effects. If white troopers are more biased against Hispanic motorists, then I β_1 and β_2 should be negative.

I report my results in Table 17. I find that Hispanic troopers are 3.3% more likely to mask Hispanic motorists compared to white troopers. This is a surprising result compared to past literature, notably Goncalves and Mello (2017), which found lower levels of trooper bias for own-race stops. I find that black troopers are less likely to mask when compared to white troopers, but the coefficient is insignificant. I find no significant differences in masking behavior across trooper race for black motorists, but the coefficients are negative.

Next, I test if employment outcomes, such as salary and experience, are related to bias where experience is measured using the hire year of the trooper. This also provides insight in how DPS was responding to bias during that time. Were biased troopers being promoted more than unbiased troopers? Do biased troopers get paid more? Does bias increase with experience?

I show my results in Table **??**. I find no correlations of trooper salary and experience to Hispanic bias. Trooper salary increases in a systematic manner with years of experience conditional on staying within the same rank. The only employment characteristic correlated to bias is trooper rank. I find that probationary troopers are -0.21 standard deviations lower in bias compared to troopers. I also find that

Electronic copy available at: https://ssrn.com/abstract=3357063

troopers who increase in rank to corporal are also less biased, but this coefficient is not significant. It appears that troopers on probation perform less biased, but once promoted to full trooper, behave in a more biased way. But, the lack of significance and small size of the coefficients for higher ranks indicates biased troopers do not get promoted more.

To examine the relationship between trooper bias and trooper's career across time, I divide the trooper's career into two sections: pre-2013, and 2014-2015. This has a few advantages; first, I do not have trooper employment data prior to 2013 so 2013 is the earliest year I can use. Second, the measure of bias has high variance since it's measured on the differential masking behavior across search outcomes. Therefore, officers with few searches have high variance. By dividing the trooper's career into two sections rather than by year, my estimate of bias is more efficient and more accurate. Lastly, with the panel-like structure, I can test if changes in employment outcomes are related to bias, specifically outcomes such as increasing in rank, moving cities, and leaving the force. Moving cities is a proxy for salary increases since troopers salary increases systematically. Rather than a salary increase, a trooper can be compensated for good behavior by being stationed at a preferred city.

I show my results in Table 19. I regress the likelihood of leaving the force, moving cities, and increasing rank on the standardized measure of bias from the first half of the trooper's career including controls for trooper's rank before 2014 and for their work city. I find that the probability of leaving the force increases by 2% for every 1 standard deviation increase in Hispanic bias. I find no relationship between an officer's measure of Hispanic bias prior to 2014 has no effect on increasing in officer's rank or moving cities in 2014 and 2015.

I also break column 3 from Table 19 into each rank. I regress the probability of increasing rank for each rank of trooper interacted with the level of bias from prior to 2014. I find no evidence that more biased troopers are more or less likely to be promoted regardless of rank.

While understanding the effect of bias on employment outcomes is important, I am also interested in the effect of bias on other measures of trooper behavior such as complaints. Using complaint data from DPS from 2013 to 2015, I test whether measures of bias is related to civilian complaint. The complaint data from DPS is suppressed for cases that were unsustained, resulted in blah blah, or for blank.

The results in Table 21 show that biased troopers

5 Conclusion

In this paper, I use the cheating behavior of troopers to uncover their taste towards searching driver's by motorist race. I develop a new statistical model to use this cheating behavior to measure trooper's

Electronic copy available at: https://ssrn.com/abstract=3357063

racial bias. Unlike past tests of racial bias, my test can measure the magnitude of bias at the individual trooper level. I test my model using a rich data set of Texas highway stop data from 2010 - 2015 merged with Texas Highway Patrol employee administrative data. During this time period, Texas troopers were masking motorist race of non-White motorists as White for certain stops. By comparing these search and search success rates across motorist race, trooper race, observed and corrected, masked and unmasked, I am able to develop a comprehensive test for racial bias.

In my results, I find that black troopers are unbiased against non-white motorists and white and Hispanic troopers are biased against Hispanic motorists. White troopers are also biased against black motorists. By masking many Hispanic motorists as white, troopers appear less successful in searching white motorists and are able to avoid being labeled biased.

Using my measure of bias, I also examine the role of institutions on officer level of bias. I find that moving to a county with one standard deviation higher in average bias increases the officer's own bias by one standard deviation, regardless of the county being located by the Mexican border. This implies that even good troopers may behave badly as a result of environment rather than from their own tastes and preferences, highlighting the importance of institutions and social norms on officer bias. Further research is needed to examine whether changing institutions may be more effective and efficient in curbing biased behavior compared to officer-level interventions.

Masking may not just be present in law enforcement. In any sort of scenario where racial profiling is illegal, this may induce agents to mask the race of the biased group to appear less biased. For example in mortgage lending, mortgage lenders may mask the race of applicants to appear less biased. This masking may not be limited to just race but is also easily extended to other observable characteristics such as income or educational level.

Electronic copy available at: https://ssrn.com/abstract=3357063

6 Appendix

References

Alesina, A. and E. L. Ferrara (2014). A test of racial bias in capital sentencing. *American Economic Review*.

Anbarci, N. and J. Lee (2014). Detecting racial bias in speed discounting: evidence from speeding tickets in Boston. *International Review of Law and Economics*.

Antonovics, K. and B. G. Knight (2009). A new look at racial profiling: evidence from the Boston police department. *The Review of Economics and Statistics*.

Anwar, S. and H. Fang (2006). An alternative test of racial prejudice in motor vehicle searches: theory and evidence. *American Economic Review*.

Anwar, S. and H. Fang (2015). Testing for racial prejudice in the parole board release process: theory and evidence. *Journal of Legal Studies*.

Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *Quarterly Journal of Economics*.

Collister, B. (2015a). DPS troopers getting race right after KXAN investigation. *KXAN*.

Collister, B. (2015b). Texas troopers ticketing Hispanic drivers as white. *KXAN*.

Coviello, D. and N. Persico (2013). An economic analysis of black-white disparities in NYPD's stop and frisk program. Working Paper 18803, NBER, https://www.nber.org/papers/w18803.

DPS. *When stopped by law enforcement.* Texas DPS.

Fiscella, K. and A. M. Fremont (2006). Use of geocoding and surname analysis to estimate race and ethnicity. *HSR: Health Services Research*.

Goncalves, F. and S. Mello (2017). A few bad apples? Racial bias in policing. Working Paper 608, IRS Working Papers, http://arks.princeton.edu/ark:/88435/dsp01z890rw746.

Harris, D. A. (1999). Driving while black: racial profiling on our nation's highways. *American Civil Liberties Union Special Report*.

Horrace, W. C. and S. M. Rohlin (2016). How dark is dark? Bright lights, big city, racial profiling. *The Review of Economics and Statistics*.

Electronic copy available at: https://ssrn.com/abstract=3357063

Jacob, B. A. and S. D. Levitt (2003). Rotten apples: an investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*.

Kalinowski, J., S. L. Ross, and M. B. Ross (2017). Endogenous driving behavior in veil of darkness tests for racial profiling. Working Papers 2017-017, Human Capital and Economic Opportunity Working Group, https://ideas.repec.org/p/hka/wpaper/2017-017.html.

Knowles, J., N. Persico, and P. Todd (2001). Racial bias in motor vehicle searches: theory and evidence. *Journal of Political Economy*.

Knox, D., W. Lowe, and J. Mummolo (2019). The bias is built n: how administrative records mask racially biased policing. Technical report, SSRN.

Lind, D. (2015). The FBI is trying to get better data on police killings. Here's what we know now. *Vox*.

Oyeniyi, D. (2015). State troopers will now just ask drivers their race. *TexasMonthly*.

Persico, N. and P. E. Todd (2005). Passenger profiling, imperfect screening, and airport security. *American Economic Association Papers and Proceedings*.

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

Roland G. Fryer, J. (forthcoming). An empirical analysis of racial differences in police use of force. Working Paper 22399, NBER, https://www.nber.org/papers/w22399.

Shayo, M. and A. Zussman (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics*.

TxDPS (2018). Complaint investigation and resolution.

West, J. (2018). Racial bias in police investigations.

Electronic copy available at: https://ssrn.com/abstract=3357063

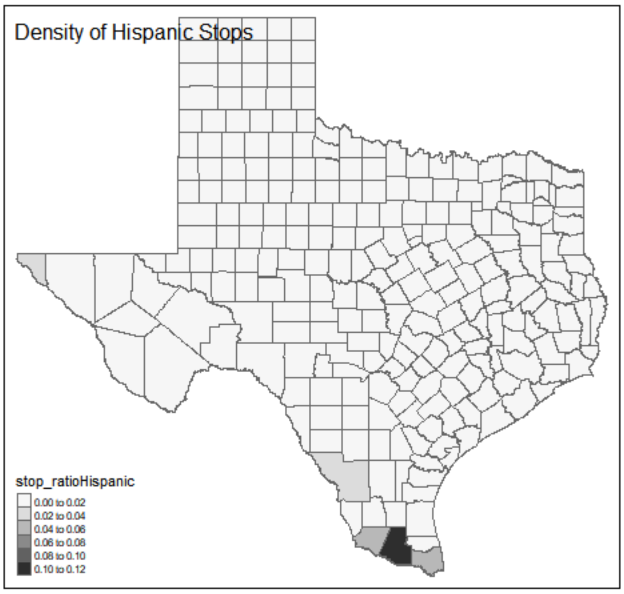


Figure 1: Hispanic Motorist Stop Density by County

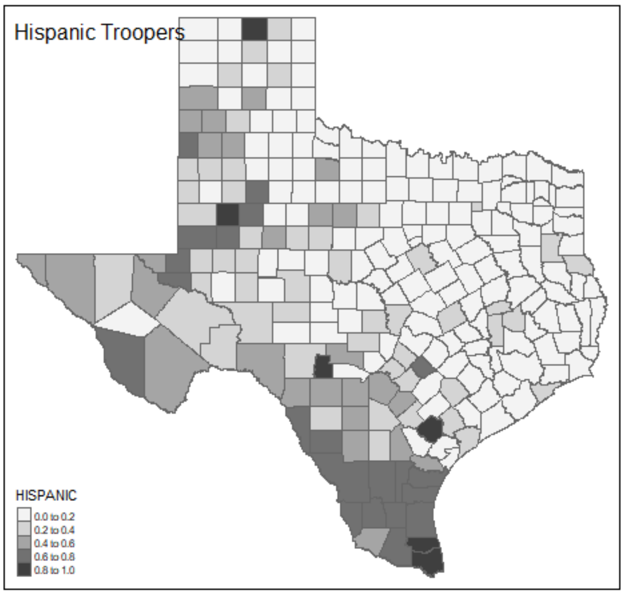


Figure 2: Hispanic Trooper Density by County

Figure 3: Black Stops Over Time - Observed and Corrected Races

Electronic copy available at: https://ssrn.com/abstract=3357063

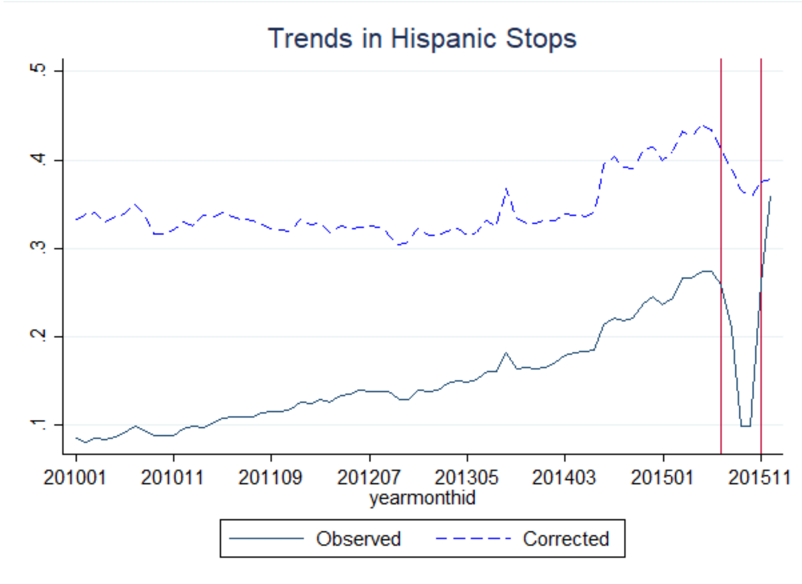


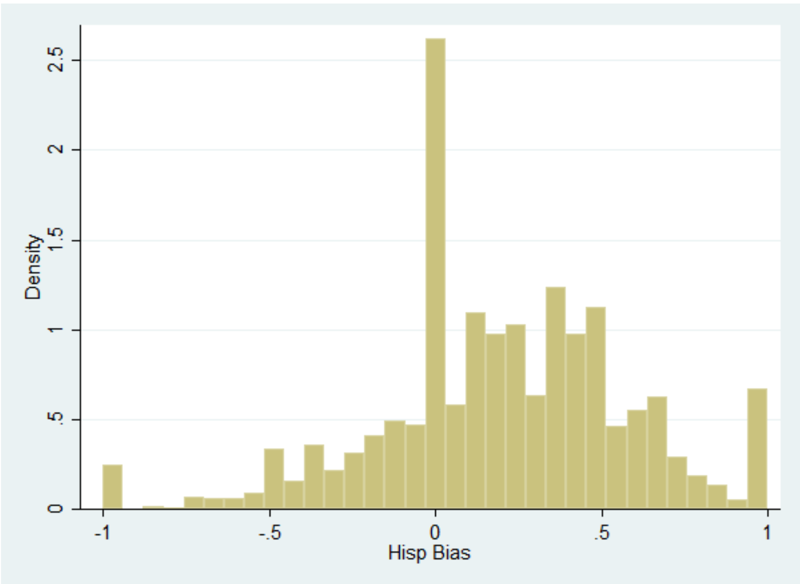
Figure 4: Hispanic Stops Over Time - Observed and Corrected Races

Figure 5: White Stops Over Time - Observed and Corrected Races

Figure 6: Example of Ticket

Figure 7: Trends in Masking across Time

Figure 8: Histogram of Hispanic Bias Measure



Electronic copy available at: https://ssrn.com/abstract=3357063

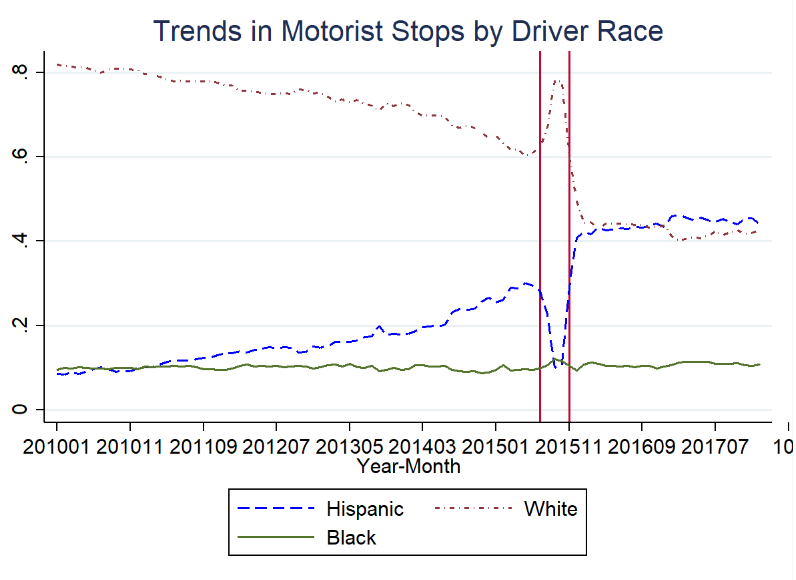
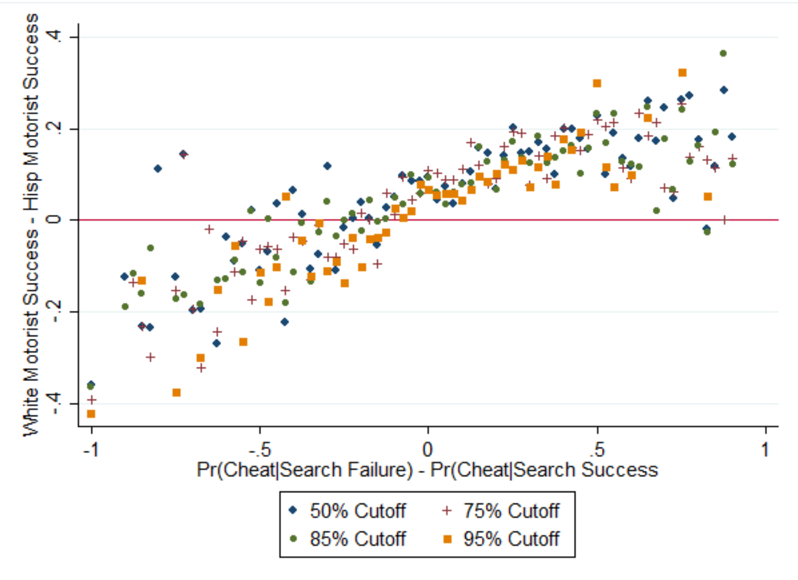


Figure 9: Stops Over Time - Observed and Corrected Races

Figure 10: KPT and Masking Measure of Bias with different thresholds



Electronic copy available at: https://ssrn.com/abstract=3357063

Figure 11: Natural Experiment - Hispanic Motorists

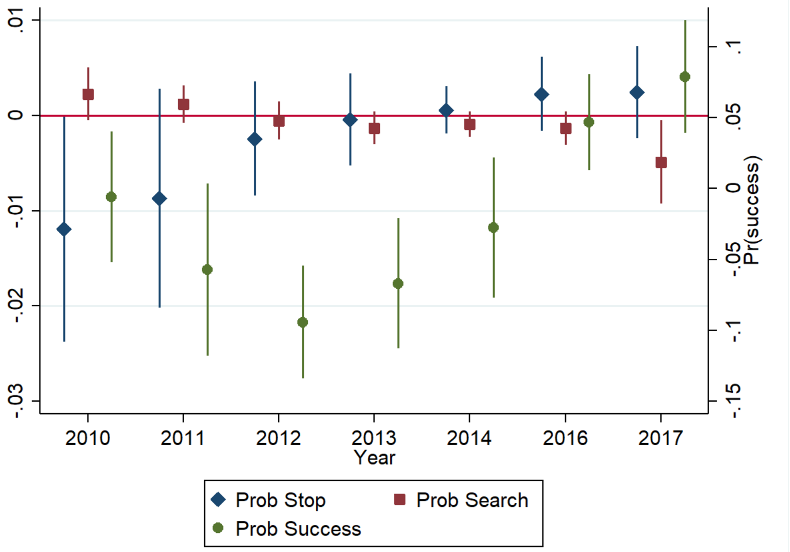
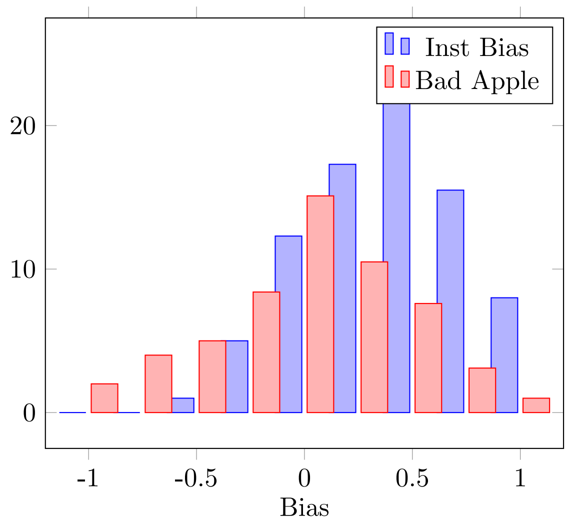


Figure 12: Examples of Institutional Bias and Bad Apples



Electronic copy available at: https://ssrn.com/abstract=3357063

Figure 13: Institutional Bias Map

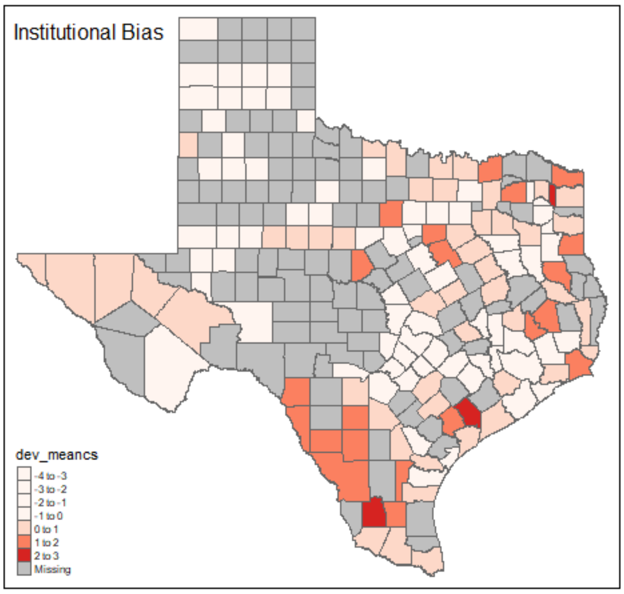
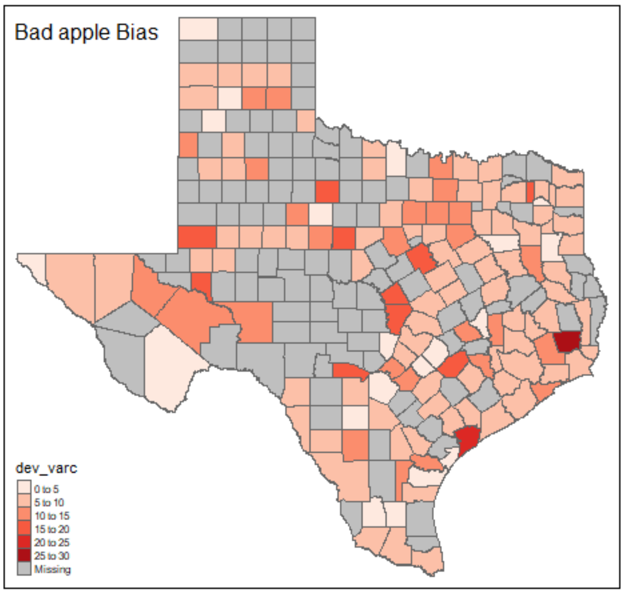


Figure 14: Bad Apple Map



Electronic copy available at: https://ssrn.com/abstract=3357063

Figure 15: Institutional Biased counties across time - Hispanic Motorists

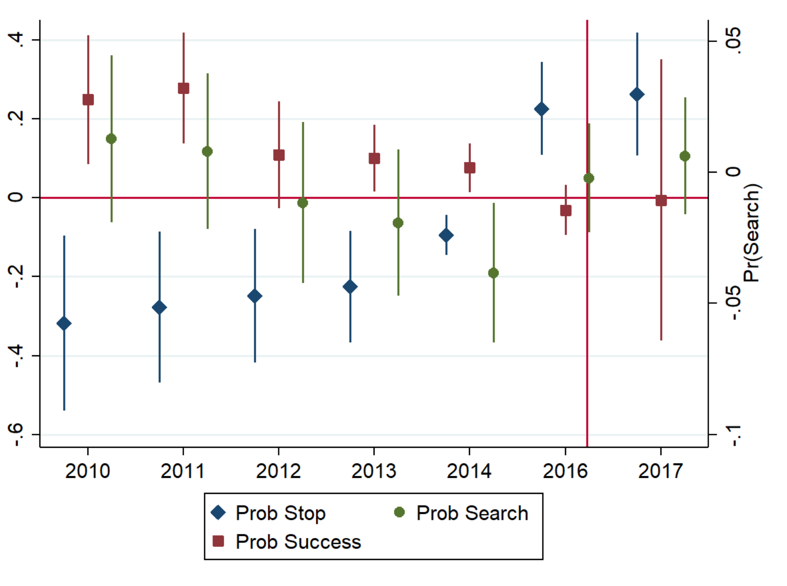
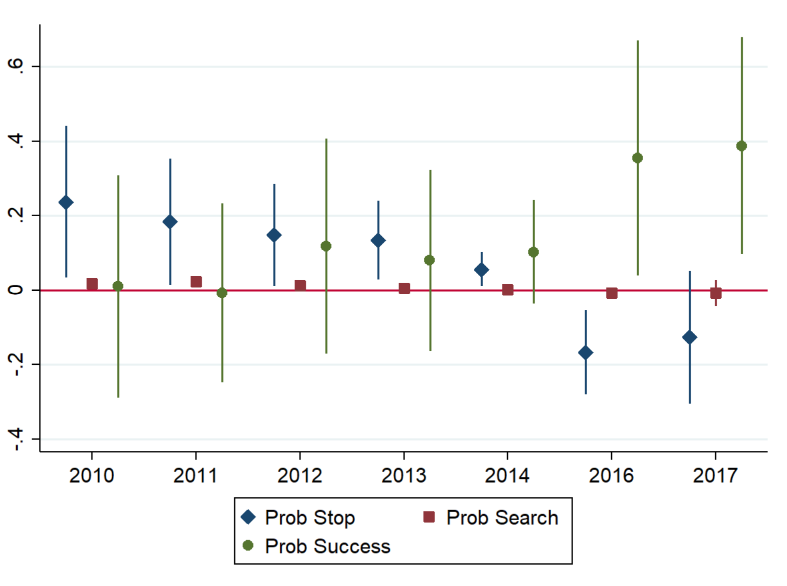


Figure 16: Institutional Biased counties across time - White Motorists



Electronic copy available at: https://ssrn.com/abstract=3357063

Figure 17: Bad apple counties across time

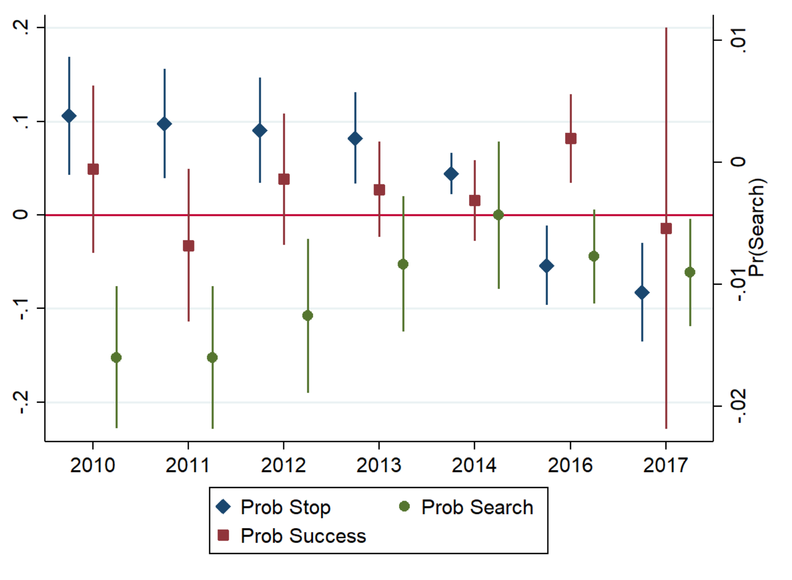
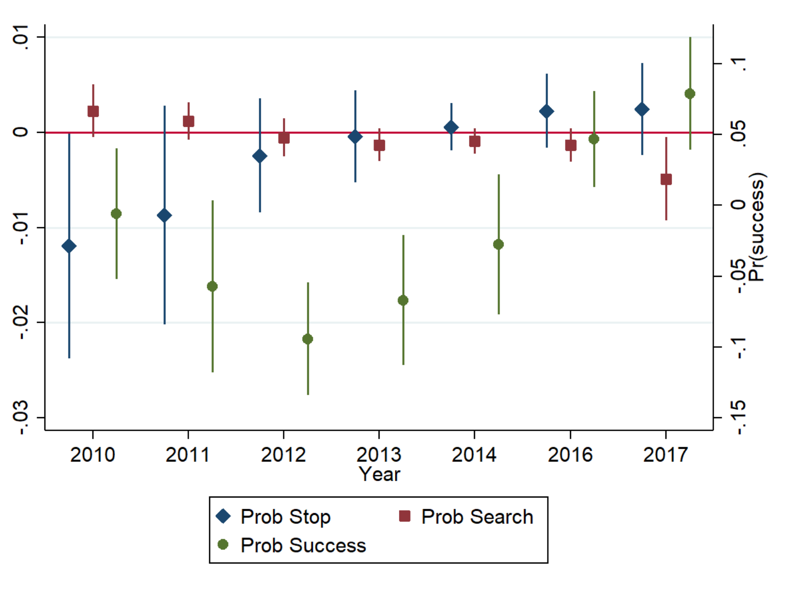


Figure 18: Natural Experiment - Hispanic Motorists



Electronic copy available at: https://ssrn.com/abstract=3357063

Figure 19: Natural Experiment - White Motorists

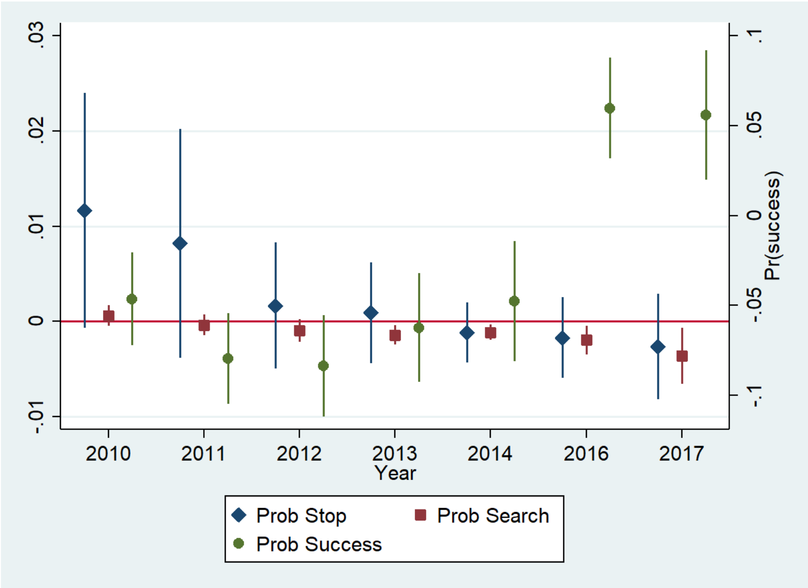


Table 1: Aggregate Statistics with Masking

	Stops			Searches		
	True	Observed	Δ	True	Observed	Δ
Driver Race						
Black	.108	.106	.002	.201	.199	.002
	(.311)	(.308)	(3.9)	(.401)	(.399)	(.6)
	734358	720449	13909	28323	28005	318
Hispanic	.341	.136	.205	.402	.17	.232
	(.474)	(.343)	(432.4)	(.49)	(.375)	(72.9)
	2311253	922138	1389115	56541	23871	32670
White	.549	.723	-.174	.396	.586	-.19
	(.498)	(.447)	(-531.4)	(.489)	(.493)	(-70.6)
	3728593	4908821	-1180228	55720	82464	-26744

Standard deviations in parentheses except for Columns 3 and 6 where the Z-stat is shown. Unweighted means are shown

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 2: Mean of Variables Related to Drivers

Driver Characteristics	All	Searches Only	Δ
Black	.108	.201	-.093
	(.311)	(.401)	(.001)
Hispanic	.341	.402	-.061
	(.474)	(.49)	(.001)
White	.549	.396	.153
	(.498)	(.489)	(.001)
Midnight	.088	.131	-.043
	(.283)	(.337)	(.001)
OwnerDriver	.201	.142	.059
	(.401)	(.35)	(.001)
TexasDriver	.896	.848	.048
	(.305)	(.359)	(.001)
OldCar	.307	.427	-.12
	(.461)	(.495)	(.001)
NewCar	.331	.177	.154
	(.471)	(.382)	(.001)
LuxuryCar	.083	.101	-.018
	(.276)	(.301)	(.001)

Standard deviations are in parentheses. Unweighted means are shown.

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 3: Mean of Variables Related to Troopers

Troopers' Characteristics	All	Searches	Masked Stops (Non-Border only)	
Black	.086	.012	.182	.165
	(.281)	(.107)	(.386)	(.371)
Hispanic	.3	.016	.341	.255
	(.458)	(.124)	(.474)	(.436)
White	.604	.02	.189	.171
	(.489)	(.139)	(.392)	(.376)
Male	.943	.018	.228	.185
	(.232)	(.134)	(.42)	(.389)
Captain	.003	.006	.041	.041
	(.059)	(.077)	(.199)	(.199)
Lieutenant	.012	.033	.317	.229
	(.108)	(.179)	(.465)	(.42)
Sergeant	.106	.024	.298	.228
	(.307)	(.153)	(.458)	(.42)
Corporal	.095	.017	.229	.184
	(.294)	(.128)	(.42)	(.387)
Trooper	.702	.018	.224	.183
	(.457)	(.132)	(.417)	(.387)
ProbTrooper	.081	.016	.249	.199
	(.273)	(.127)	(.433)	(.399)
NoRank	0	.02	.276	.224
	(.02)	(.139)	(.447)	(.417)
Total Observations	2605			

Only merged observations are shown. Trooper rank uses the highest rank the trooper obtained during 2010 - 2015

Table 4: Search Success Rates across Driver's Race

	Search Success Rate		
	(1)	(2)	(3))
	Observed	Corrected	Δ
Driver Race			
Black	.422	.421	.001
	(.494)	(.494)	(.004)
	27999	28317	-318
Hispanic	.307	.297	.01
	(.461)	(.457)	(.004)
	23868	56530	-32662
White	.421	.482	-.061
	(.494)	(.5)	(.003)
	82451	55715	26736

Unweighted means are shown using the resampled data. Standard deviations are in the parentheses. Columns 2 and 3 exclude border counties. Columns 1-3 use data from January 2010 - June 2015.

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 5: Difference in Mask Rate by Search Success

Driver Race	$Pr(Mask \mid Failure)$	$Pr(Mask \mid Success)$	Δ
Black	.012	.009	.003
	(.11)	(.094)	(.001)
Hispanic	.589	.566	.023
	(.492)	(.496)	(.005)

Unweighted means are shown using non-border counties only from January 2010 to June 2015. Standard deviations are in parantheses

Table 6: Bias and Search

	(1)	(2)
	$Pr(Search)$	$Pr(SearchSuccess)$
Hisp Bias	-0.001^{**}	-0.147^{***}
	(0.000)	(0.007)
I(Black=1)*Hisp Bias	-0.001	0.013^{*}
	(0.001)	(0.007)
I(Hispanic=1)*Hisp Bias	0.002^{***}	-0.070^{***}
	(0.000)	(0.011)
Constant	0.019^{***}	0.436^{***}
	(0.006)	(0.047)
Observations	6101499	127970

Hisp Bias is the normalized measure of Hispanic bias for each trooper. The regression includes county FE, year FE, and county x year FE along with driver race FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. $I(Black = 1)$ and $I(Hispanic = 1)$ are indicator variables for the driver's race. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 7: Search Decision for Biased Stops

	(1)	(2)	(3)	(4)
	Black	*Black*	*Hisp*	*Hisp*
I(Consent = 1)	0.008^{***}		0.003	
	(0.002)		(0.008)	
Luxury Vehicle		-0.008^{***}		-0.005
		(0.003)		(0.012)
Owner is Driver		-0.003		0.063^{***}
		(0.003)		(0.008)
Midnight Stop		0.002		-0.013^{*}
		(0.003)		(0.008)
In State		0.000		0.044^{***}
		(.)		(0.013)
Constant	0.008^{***}	0.016^{***}	0.569^{***}	0.619^{***}
	(0.001)	(0.001)	(0.007)	(0.012)
Observations	12843	8666	42811	24837

Dependent variable is an indicator variable equal to 1 if the stop is a biased stop and where a biased stop is defined as a failed search where the driver's race was misrecorded. The regression includes county FE, year FE, and county x year FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. Consent is equal to 1 if the motorist consented to search and equal to 0 if the search occurred due to probable cause. Midnight Stop is an indicator variable equal to 1 if the stop was conducted between 10 PM and 6 AM. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 8: Behavior After Masking - Hispanic motorists

	(1)	(2)	(3)
	$Pr(stop)$	$Pr(Search)$	$Pr(Success)$
Hisp bias	-0.005^{***}	-0.000	-0.120^{***}
	(0.002)	(0.001)	(0.014)
Hisp Bias x I(Year = 2010)	-0.012^{**}	0.002	-0.006
	(0.006)	(0.001)	(0.023)
Hisp Bias x I(Year = 2011)	-0.009	0.001	-0.057^{*}
	(0.006)	(0.001)	(0.031)
Hisp Bias x I(Year = 2012)	-0.002	-0.001	-0.094^{***}
	(0.003)	(0.001)	(0.020)
Hisp Bias x I(Year = 2013)	-0.000	-0.001	-0.067^{***}
	(0.002)	(0.001)	(0.023)
Hisp Bias x I(Year = 2014)	0.001	-0.001	-0.028
	(0.001)	(0.001)	(0.025)
Hisp Bias x I(Year = 2016)	0.002	-0.001	0.047^{***}
	(0.002)	(0.001)	(0.017)
Hisp Bias x I(Year = 2017)	0.002	-0.005^{**}	0.079^{***}
	(0.002)	(0.002)	(0.020)
Constant	0.205^{***}	0.035^{***}	0.369^{***}
	(0.000)	(0.000)	(0.001)
Observations	8939205	1829851	64313

Dependent variable is an indicator variable equal to 1 if the stop is a biased stop and where a biased stop is defined as a failed search where the driver's race was misrecorded. The regression includes county FE, year FE, and county x year FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. Consent is equal to 1 if the motorist consented to search and equal to 0 if the search occurred due to probable cause. Midnight Stop is an indicator variable equal to 1 if the stop was conducted between 10 PM and 6 AM. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 9: Behavior After Masking - White motorists

	(1)	(2)	(3)
	$Pr(stop)$	$Pr(Search)$	$Pr(Success)$
Hisp bias	0.006^{***}	-0.000	-0.106^{***}
	(0.002)	(0.000)	(0.012)
Hisp Bias x I(Year = 2010)	0.012^{*}	0.001	-0.046^{***}
	(0.006)	(0.001)	(0.013)
Hisp Bias x I(Year = 2011)	0.008	-0.000	-0.080^{***}
	(0.006)	(0.001)	(0.013)
Hisp Bias x I(Year = 2012)	0.002	-0.001	-0.084^{***}
	(0.003)	(0.001)	(0.014)
Hisp Bias x I(Year = 2013)	0.001	-0.001^{***}	-0.062^{***}
	(0.003)	(0.001)	(0.015)
Hisp Bias x I(Year = 2014)	-0.001	-0.001^{***}	-0.047^{***}
	(0.002)	(0.000)	(0.017)
Hisp Bias x I(Year = 2016)	-0.002	-0.002^{**}	0.060^{***}
	(0.002)	(0.001)	(0.014)
Hisp Bias x I(Year = 2017)	-0.003	-0.004^{**}	0.056^{***}
	(0.003)	(0.001)	(0.018)
Constant	0.676^{***}	0.020^{***}	0.454^{***}
	(0.000)	(0.000)	(0.001)
Observations	8939205	6046709	119509

Dependent variable is an indicator variable equal to 1 if the stop is a biased stop and where a biased stop is defined as a failed search where the driver's race was misrecorded. The regression includes county FE, year FE, and county x year FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. Consent is equal to 1 if the motorist consented to search and equal to 0 if the search occurred due to probable cause. Midnight Stop is an indicator variable equal to 1 if the stop was conducted between 10 PM and 6 AM. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 10: Bias and Social Norms

	(1) Inst Bias	(2) Coef Var
% Hisp	-0.36	-0.74
	(0.23)	(0.52)
% Black	-0.24	-0.80*
	(0.20)	(0.44)
% no health ins	0.29	1.03
	(0.49)	(1.16)
% HS diploma	0.04	1.09
	(0.45)	(1.22)
Median HH inc (10000s)	-0.00	-0.03
	(0.02)	(0.04)
% Employed	-1.32***	-4.40***
	(0.50)	(1.55)
% older than 16	-0.65	-1.42
	(0.79)	(1.88)
Population (100000s)	-0.01***	-0.02***
	(0.00)	(0.01)
Border County	0.19**	0.46***
	(0.08)	(0.16)
Constant	1.91***	5.68***
	(0.67)	(1.74)
Observations	131	131
R^2	0.240	0.332
F	7.18	5.59

Inst Bias is equal to the difference in mean of Hispanic bias between the county and all of Texas. Bad Apples is the ratio of standard deviations of Hispanic bias of the county and Texas. Coef Var is equal to the standard deviation of Hispanic bias of the county divided by the mean of the county. Ratio of Var is the ratio of variation of Hispanic bias of the county and Texas. County level variables are from the 2010 - 2015 ACS. Counties with less than 4 stops annually were excluded.

Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 11: Bias and Social Norms

	(1) Δ Hisp Bias	(2) Δ Hisp Bias	(3) Δ Hisp Bias
Δ Mean Bias	0.48***	0.34***	0.30***
	(0.05)	(0.06)	(0.05)
Border-NonBorder		-0.27***	
		(0.06)	
Δ Mean Bias*Border-NonBorder		0.97***	
		(0.03)	
NonBorder-Border			0.09
			(0.06)
Δ Mean Bias*NonBorder-Border			0.92***
			(0.02)
Constant	0.05	0.09*	0.09*
	(0.04)	(0.05)	(0.05)
Observations	572	462	536
R^2	0.125	0.158	0.328
F	78.99	522.46	1251.11

Dependent variable is the difference in officer-county bias for each officer's two most searched counties. Only includes officers who have conducted at least 4 searches in two counties.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 12: Bias and Social Norms

	(1) Δ Hisp Bias	(2) Δ Hisp Bias	(3) Δ Hisp Bias
Δ CoefVar(Bias)	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Border-NonBorder		0.15**	
		(0.07)	
Δ CoefVar(Bias)*Border-NonBorder		0.13***	
		(0.04)	
NonBorder-Border			-0.22***
			(0.05)
Δ CoefVar(Bias)*NonBorder to Border			0.00
			(0.00)
Constant	-0.03*	0.00	0.00
	(0.02)	(0.02)	(0.02)
Observations	378	300	353
R^2	0.002	0.024	0.060
F	0.98	6.59	10.02

Dependent variable is the coefficient of variation in officer-county bias for each officer's two most searched counties. Only includes officers who have conducted at least 4 searches in two counties.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 13: Bias and Social Norms-Hispanic Motorists

	(1) $Pr(stop)$	(2) $Pr(Search)$	(3) $Pr(Success)$
Inst Bias	0.080^{***}	-0.003^{**}	-0.144^{***}
	(0.027)	(0.001)	(0.020)
Inst Bias x I(Year = 2010)	-0.057^{***}	0.005^{**}	0.027
	(0.020)	(0.002)	(0.019)
Inst Bias x I(Year = 2011)	-0.050^{***}	0.006^{***}	0.021
	(0.017)	(0.002)	(0.018)
Inst Bias x I(Year = 2012)	-0.045^{***}	0.001	-0.002
	(0.015)	(0.002)	(0.019)
Inst Bias x I(Year = 2013)	-0.041^{***}	0.001	-0.011
	(0.013)	(0.001)	(0.017)
Inst Bias x I(Year = 2014)	-0.017^{***}	0.000	-0.034^{**}
	(0.005)	(0.001)	(0.016)
Inst Bias x I(Year = 2016)	0.041^{***}	-0.003^{***}	0.009
	(0.011)	(0.001)	(0.013)
Inst Bias x I(Year = 2017)	0.047^{***}	-0.002	0.019
	(0.014)	(0.005)	(0.013)
Constant	0.250^{***}	0.018^{***}	0.403^{***}
	(0.028)	(0.001)	(0.021)
Observations	9867447	2496954	89329

Dependent variable is the difference in officer-county bias for each officer's two most searched counties. Only includes officers who have conducted at least 4 searches in two counties.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 14: Bias and Social Norms-White Motorists

	(1) $Pr(stop)$	(2) $Pr(Search)$	(3) $Pr(Success)$
Inst Bias	-0.050^{**}	-0.002^{**}	-0.113^{***}
	(0.023)	(0.001)	(0.028)
Inst Bias x I(Year = 2010)	0.043^{**}	0.003^{***}	0.002
	(0.019)	(0.001)	(0.027)
Inst Bias x I(Year = 2011)	0.033^{**}	0.004^{***}	-0.001
	(0.016)	(0.001)	(0.022)
Inst Bias x I(Year = 2012)	0.027^{**}	0.002^{**}	0.021
	(0.013)	(0.001)	(0.026)
Inst Bias x I(Year = 2013)	0.024^{**}	0.001	0.014
	(0.010)	(0.001)	(0.022)
Inst Bias x I(Year = 2014)	0.010^{**}	0.000	0.019
	(0.004)	(0.001)	(0.013)
Inst Bias x I(Year = 2016)	-0.030^{***}	-0.001	0.064^{**}
	(0.010)	(0.001)	(0.029)
Inst Bias x I(Year = 2017)	-0.023	-0.002	0.070^{***}
	(0.016)	(0.003)	(0.027)
Constant	0.631^{***}	0.012^{***}	0.503^{***}
	(0.023)	(0.001)	(0.023)
Observations	9867447	6153056	118407

Dependent variable is the difference in officer-county bias for each officer's two most searched counties. Only includes officers who have conducted at least 4 searches in two counties.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 15: Bias and Social Norms-Hispanic Motorists

	(1)	(2)	(3)
	$Pr(stop)$	$Pr(Search)$	$Pr(Success)$
dev_sdc	-0.159^{***}	0.003	0.179^{***}
	(0.050)	(0.002)	(0.039)
Bad Apple x I(Year = 2010)	0.106^{***}	-0.001	-0.152^{***}
	(0.032)	(0.003)	(0.038)
Bad Apple x I(Year = 2011)	0.098^{***}	-0.007^{**}	-0.152^{***}
	(0.029)	(0.003)	(0.039)
Bad Apple x I(Year = 2012)	0.090^{***}	-0.001	-0.108^{**}
	(0.029)	(0.003)	(0.042)
Bad Apple x I(Year = 2013)	0.082^{***}	-0.002	-0.052
	(0.025)	(0.002)	(0.037)
Bad Apple x I(Year = 2014)	0.044^{***}	-0.003^{*}	-0.000
	(0.011)	(0.002)	(0.040)
Bad Apple x I(Year = 2016)	-0.054^{**}	0.002	-0.044^{*}
	(0.021)	(0.002)	(0.025)
Bad Apple x I(Year = 2017)	-0.082^{***}	-0.005	-0.062^{**}
	(0.027)	(0.008)	(0.029)
Constant	0.705^{***}	0.011^{*}	-0.111
	(0.165)	(0.006)	(0.122)
Observations	9867447	2496954	89329

Dependent variable is the difference in officer-county bias for each officer's two most searched counties. Only includes officers who have conducted at least 4 searches in two counties.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 16: Bias and Social Norms-White Motorists

	(1)	(2)	(3)
	$Pr(stop)$	$Pr(Search)$	$Pr(Success)$
Bad Apple	0.114^{***}	-0.000	0.144^{***}
	(0.037)	(0.001)	(0.043)
Bad Apple x I(Year = 2010)	-0.085^{***}	-0.003^{***}	-0.094^{**}
	(0.026)	(0.001)	(0.040)
Bad Apple x I(Year = 2011)	-0.073^{***}	-0.005^{***}	-0.068^{**}
	(0.023)	(0.001)	(0.031)
Bad Apple x I(Year = 2012)	-0.064^{***}	-0.002^{*}	-0.109^{***}
	(0.021)	(0.001)	(0.041)
Bad Apple x I(Year = 2013)	-0.058^{***}	-0.002^{*}	-0.079^{**}
	(0.018)	(0.001)	(0.032)
Bad Apple x I(Year = 2014)	-0.033^{***}	-0.000	-0.047^{**}
	(0.009)	(0.001)	(0.019)
Bad Apple x I(Year = 2016)	0.039^{*}	-0.001	-0.087^{**}
	(0.020)	(0.002)	(0.042)
Bad Apple x I(Year = 2017)	0.058^{***}	-0.001	-0.113^{***}
	(0.022)	(0.005)	(0.038)
Constant	0.305^{**}	0.013^{***}	0.096
	(0.121)	(0.003)	(0.140)
Observations	9867447	6153056	118407

Dependent variable is the difference in officer-county bias for each officer's two most searched counties. Only includes officers who have conducted at least 4 searches in two counties.
Regression uses robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 17: Hispanic Bias and Trooper Race

	(1)	(2)
	BlackMotorists	*HispMotorists*
Hisp Troopers	−0.0000	−0.0022
	(0.0001)	(0.0027)
Black Troopers	-0.0001^{*}	-0.0106^{***}
	(0.0001)	(0.0030)
Constant	0.0003^{***}	0.0250^{***}
	(0.0000)	(0.0012)
Observations	625912	827919

Dependent variable is an indicator variable equal to 1 if the stop is a biased stop. Biased stop is equal to one for failed search where the driver's race was misrecorded and zero it was a failed search where the driver's race was correctly recorded. The regression includes county FE, year FE, and county x year FE. Standard errors are clustered at the county level. Regression uses data from January 2010 - June 2015. Hisp Troopers and Black Troopers are indicator variables for the trooper's race with white troopers being the omitted category. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 18: Hisp Bias on Labor Outcomes

	(1)	(2)	(3)
	Hisp Bias	Hisp Bias	Hisp Bias
Experience	0.001		
	(0.002)		
Salary		0.016	
		(0.014)	
Prob. Troop			-0.202^{***}
			(0.077)
Corporal			−0.035
			(0.030)
Sergeant			0.013
			(0.036)
Lieutenant			−0.011
			(0.146)
Observations	1834	1834	1834

Regression has robust standard errors show in parentheses. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063

Table 19: Hisp Bias on Labor Outcomes - Panel

	(1) $Pr(LeftForce)$	(2) $Pr(MovedCities)$	(3) $Pr(RankUp)$
Hisp bias	0.0213	-0.0081	-0.0018
	(0.0147)	(0.0136)	(0.0158)
Prob. Troop	-0.1835	0.0867	0.8358^{***}
	(0.1219)	(0.1496)	(0.1053)
Corporal	0.0536	-0.0523	0.0275
	(0.0595)	(0.0426)	(0.0568)
Sergeant	-0.0726	0.0471	-0.1034^{*}
	(0.0583)	(0.0584)	(0.0605)
Lieutenant	-0.3086^{***}	-0.0023	0.0001
	(0.0391)	(0.0429)	(0.0462)
Black Trooper	0.0091	-0.0519	-0.1216^{*}
	(0.0821)	(0.0687)	(0.0722)
Hisp Trooper	-0.0024	-0.0033	-0.0038
	(0.0468)	(0.0437)	(0.0451)
Observations	837	724	724

Dependent variable is the officer level measure of Hispanic bias. Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 20: Hisp Bias on Labor Outcomes - Transition Matrix

	(1)
Prob Trooper x Hisp Bias	-0.015
	(0.014)
Corporal x Hisp Bias	-0.005
	(0.038)
Sergeant x Hisp Bias	0.028
	(0.028)
Observations	718

Dependent variable is the probability of increasing in rank. Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Table 21: Hisp Bias on Complaints

	(1) Complained	(2) Sustained
Hisp Bias	0.007	0.007
	(0.005)	(0.005)
Constant	0.058^{***}	0.055^{***}
	(0.005)	(0.005)
Observations	2234	2234
R^2	0.001	0.001
F	1.931	1.996

Regression has robust standard errors. * $p < 0.1$; ** $p < 0.5$; *** $p < 0.01$

Electronic copy available at: https://ssrn.com/abstract=3357063