ARE CONSTITUTIONAL RIGHTS ENOUGH?
AN EMPIRICAL ASSESSMENT OF RACIAL BIAS IN POLICE STOPS

Rohit Asirvatham & Michael D. Frakes

ABSTRACT—This Article empirically tests the conventional wisdom that a permissive constitutional standard bearing on pretextual traffic stops—such as the one announced by the Supreme Court in Whren v. United States—contributes to racial disparities in traffic stops. To gain empirical traction on this question, we look to state constitutional law. In particular, we consider a natural experiment afforded by changes in the State of Washington’s rules regarding traffic stops. Following Whren, the Washington Supreme Court first took a more restrictive stance than the U.S. Supreme Court, prohibiting pretextual stops by police officers, but later reversed course and instituted a laxer standard, effectively equivalent to Whren’s.

We investigate the effect of this retreat to a Whren-like standard on the degree of racial disparities in traffic stops in Washington. For that purpose, we use a dataset of over 7 million traffic stops and employ a range of empirical techniques—including the estimation of difference-in-difference and triple-differences specifications—that are designed to isolate the effect of the change in Washington constitutional law and account for both observable and unobservable factors that may also impact racial disparities in traffic-stop rates. In particular, we employ a novel methodological approach designed to separate the effect of the change in constitutional standards in Washington from the effect of Washington’s contemporaneous legalization of recreational marijuana.

Across our deep dive into these matters, we fail to find evidence that supports the conventional wisdom that a Whren-like standard intensifies racial bias in officers’ decisions to initiate stops. On the contrary, our results suggest that constitutional standards, at best, have little to no impact on the gap between the stop rates of non-white and white drivers (or between Black and white drivers).

Racial disparities in traffic stops are an undeniable problem. And to best address this problem, we need to understand which legal tools do and do not work in regulating officer behavior. We suggest our findings may be due to certain inherent weaknesses in the way in which the relevant constitutional standards are enforced—i.e., via the exclusionary rule. To the extent an
officer’s decision to initiate a traffic stop is heavily driven by factors other than the remote possibility that any evidence obtained during a pretextual stop will be suppressed, it is unlikely that a constitutional-rights-based approach will meaningfully reduce racial disparities in traffic-stop rates. Instead, we propose several extraconstitutional approaches to this critical problem, including the use of administrative disciplinary systems that evaluate an officer’s aggregate pattern of behavior, not their behavior in individual cases—i.e., approaches designed to bolster a deterrent channel—along with the development of technologies that rely less on officer discretion in the first place.

AUTHORS—Rohit Asirvatham is a law clerk to Judge David R. Stras, United States Court of Appeals for the Eighth Circuit. Michael Frakes is the A. Kenneth Pye Professor of Law and Professor of Economics at Duke University and a Research Associate at the National Bureau of Economic Research. We are grateful for helpful comments from Thomas Frampton, Brandon Garrett, Benjamin Grunwald, Janet Moore, Kyle Rozema, Christopher Slobogin, Megan Stevenson and seminar participants at the UVA Law and Economics Colloquium and the Duke Center for Science and Justice Workshop. All views expressed in this Article are our own and do not reflect the views of Judge Stras, the Eighth Circuit, or the federal judiciary.
INTRODUCTION

By July 6, 2016, getting pulled over was routine for Philando Castile. The Black thirty-two-year-old school cafeteria worker had already been stopped at least fifty-two times in the Minneapolis–St. Paul area for reasons including not wearing a seat belt, speeding, and driving at night with an unlit license plate. This time, the officer told Castile he was being pulled over for driving with a broken taillight. But perhaps that was not entirely true. Before

---


stopping him, Officer Yanez had radioed a colleague that Castile fit the description of a robbery suspect, citing Castile’s “wide-set nose.” Nonetheless, when Officer Yanez walked up to Castile’s car, he said, “the reason I pulled you over [is] your brake lights are out.” Minutes later, Philando Castile was dead, shot by Officer Yanez.

In his final minutes, Walter Scott, a fifty-year-old Black man, heard something similar to what Castile had: “The reason for the stop is your brake light is out.” Sam DuBose heard: “You don’t have a front license plate on your car.” Sandra Bland heard: “You failed to signal your lane change.”

Traffic stops open the door to violence. They are one of the most common entry points for contact between civilians and the police. And the harms that can accompany a traffic stop encompass far more than physical violence. When an individual is stopped, her day is put on hold by the officer’s authority as an arm of the state. When the officer walks up and looks inside the car’s window, her privacy is invaded. Even the most routine stop can cause apprehension or fear, for some. And a resulting ticket or fine can have devastating effects on the driver. For example, drivers who cannot afford to pay the fine often lose their license.

---

4 CBS News, Police Dashcam Video Released in Fatal Shooting of Philando Castile, YOUTUBE (June 20, 2017), https://www.youtube.com/watch?v=9Y7sgZZQ7pw [https://perma.cc/M95B-4LHC].


7 WCPO 9, Full Video: Police Officer Ray Tensing Shoots Sam DuBose During Traffic Stop, YOUTUBE (July 29, 2015), https://www.youtube.com/watch?v=kYINt6uNjA0 [https://perma.cc/GXJ5-VV92].


10 SeeALEX BENDER, STEPHAN BINGHAM, MARI CASTALDI, ELISA DELLA PIANA, MEREDITH DESAUTELS, MICHAEL HERALD, ENDRIA RICHARDSON, JESSE STOUT & THERESA ZHAN, NOT JUST A

11 Id. at 136.

12 SeeALEX BENDER, STEPHAN BINGHAM, MARI CASTALDI, ELISA DELLA PIANA, MEREDITH DESAUTELS, MICHAEL HERALD, ENDRIA RICHARDSON, JESSE STOUT & THERESA ZHAN, NOT JUST A

---

1484
a license to work will lose their jobs.\textsuperscript{15} And that in turn makes it harder for them to pay their fines and have their licenses reinstated.\textsuperscript{16} The stakes for ensuring evenhanded traffic enforcement are high. But traffic enforcement is not evenhanded.\textsuperscript{17} The evidence is clear: Black drivers are more likely to be stopped than white drivers, even when accounting for the benchmarking problem.\textsuperscript{18}

A critical question is how to reduce this racial bias. We must figure out which avenues of attack are likely to be fruitful and which are likely to fail. A dominant view in the literature is that a robust Fourth Amendment right against pretextual stops is one such promising avenue. Our overarching goal in this Article is to assess the merits of this conventional wisdom. Ultimately, our analysis demonstrates that this dominant view—while understandable at first blush—simply lacks empirical support.

To be sure, figuring out how to regulate traffic stops presents a knotty puzzle. Given the sheer number of traffic rules, most people commit some sort of technical violation almost every time they drive. Most would agree it is not a good idea to stop people for literally every traffic violation. That would be incredibly costly, without a sufficient corresponding marginal increase in safety. But most would also agree that it is dangerous to give drivers carte blanche to violate traffic laws. That too would be costly—in


\textsuperscript{16} Id. at 6–7 (“Data shows that a valid driver’s license is a more accurate predictor of sustained employment than a General Educational Development (GED) diploma... A New Jersey study found that 42\% of people whose driver’s licenses were suspended lost their jobs as a result of the suspension.”).

\textsuperscript{17} Id. at 6.

\textsuperscript{18} See Harris, supra note 12, at 141 (“[W]e see in study after study on traffic stops and stop-and-frisks: Police stop African-Americans and Latinos more often than whites, even though stops of whites yield contraband or arrests or summonses more often.”); Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff & Sharad Goel, A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States, 4 NATURE HUM. BEHAV. 736, 740–41 (2020).
terms of safety. So the dilemma is that while we do want officers to stop some people who violate a traffic law, we do not want an officer’s decision about whether to stop a given driver to be based on bad reasons. Two such bad reasons are (1) an unsubstantiated hunch of criminal activity and (2) race. Allowing for either of these reasons to serve as the basis for a stop could increase the likelihood of racial bias in stops because a hunch might be influenced by the suspect’s race. Thus, the puzzle is how to protect drivers from being stopped on those bad grounds when, because everyone commits traffic violations, those bad grounds will almost always appear alongside a permissible reason for a stop. How do we know whether a driver was stopped because he was Black or because he was driving 7 miles per hour over the speed limit?

The Supreme Court could have said that pretextual stops—stops in which there was a traffic violation but an impermissible reason was the real motivation behind the stop—violate the Fourth Amendment. Instead, in Whren v. United States, the Court held that as long as there is a traffic violation, the stop does not violate the Fourth Amendment, regardless of the officer’s real motivations. Understandably, the decision was met with a flood of scholarly criticism taking the position that, by failing to provide a Fourth Amendment right against pretextual stops, the Court exacerbated the problem of racial bias in traffic stops. And many argue that, given the scope

that officers have to exhibit racial bias when permitted to conduct pretextual stops, a Fourth Amendment rule against pretextual stops would work to reduce that racial bias.22

 Nonetheless, there may be reasons to doubt this widely held view—that is, reasons to doubt whether constitutional prohibitions of pretextual stops would actually reduce racial disparities in traffic-stop rates in practice. First, we suggest that the means of enforcing these constitutional protections—i.e., the exclusionary rule—may insufficiently deter officers in the first place. After all, to the extent an officer’s decision to initiate a traffic stop is heavily driven by factors other than the remote possibility that any evidence obtained during a pretextual stop will be suppressed, it is unlikely that a constitutional-rights-based approach will meaningfully reduce racial disparities in traffic-stop rates. Second, and relatedly, the evidentiary barriers inherent in proving pretext may fundamentally limit the degree to which a stricter standard would impact case outcomes and thus regulate officer behavior.

 Altogether, despite the near consensus in the literature for addressing racial disparities in traffic-stop rates through a constitutional-rights-based approach, the theoretical case for that approach is simply ambiguous. Given this ambiguity, this inquiry comes down to an empirical exercise. And it is this exercise that is at the heart of our Article. We ask the question: Would a Fourth Amendment right against pretextual stops reduce racial disparities in the initiation of traffic stops? Our data suggest: no.

 Until recently, empirically assessing this question was very difficult. At the time of the Whren decision, we lacked good data on police stops. And even as data improved, having only post-Whren data did not tell us much because there was nothing to which to compare it. Two Washington Supreme Court decisions changed that. Disagreeing with Whren, the Washington Supreme Court first interpreted its state constitution’s Fourth Amendment analogue as prohibiting pretextual traffic stops.23 But thirteen years later, in 2012, it changed course and lowered that standard, sanctioning previously prohibited stops.24 As a result, we have a natural experiment through which to analyze the effect of a decision like Whren on racial bias in traffic stops.

 We analyze data from over 7 million traffic stops in Washington (along with nearly 2.5 million traffic stops in Colorado in our control analyses) and explore how racial disparities in those stops evolved in the period around the

---

22 Margaret M. Lawton, The Road to Whren and Beyond: Does the “Would Have” Test Work?, 57 DEPAUL L. REV. 917, 930–31 (2008) (“[M]any legal scholars have argued that use of a ‘reasonable officer’ test, also known as the ‘would have’ test, would combat the tendency to use traffic infractions as a pretext . . . and would consequently protect drivers.”).  
23 State v. Ladson, 979 P.2d 833, 842 (Wash. 1999) (en banc).  
2012 change in constitutional law in Washington. Our analysis employs a number of empirical techniques to distinguish the effect of this change in the law from a range of both observable and unobservable factors that may likewise impact racial disparities in the rate of traffic stops. One particular challenge of this nature is posed by the fact that the change in the constitutional treatment of pretextual stops in the State of Washington that forms the basis for this natural experiment coincided to the month with Washington’s legalization of recreational marijuana. To confront this challenge, we employ a novel approach that takes advantage of the fact that Colorado likewise legalized recreational marijuana within days of Washington but did not alter its Fourth Amendment-like treatment of pretextual stops. These facts allow us to incorporate a comparison of Colorado and Washington to untangle the effects of the two coinciding changes.

Taking a broad array of approaches, our analysis ultimately finds no robust evidence that the legal change in Washington had an effect on racial disparities in traffic stops. If anything, our findings suggest that Washington experienced a trend towards greater and greater racial disparities in traffic stops that predated the relevant change in constitutional rights, with the 2012 Washington Supreme Court decision that lowered the constitutional standard for pretextual stops doing nothing to further exacerbate the preexisting problem. Accordingly, this evidence contradicts the dominant assumption that Whren is a significant contributor to racial bias in traffic stops. And it thus cautions against looking to the Fourth Amendment for a solution to this critical problem. In its place, we propose that policymakers consider solutions that will more effectively deter officer behavior, including extraconstitutional approaches that discipline officer behavior based on their aggregate set of stops, as opposed to the choices that they make in individual cases (in which pretext is fundamentally hard to prove).

We acknowledge that we are not the first to analyze this Washington data and the relevant change in Washington law. Professors Stephen Rushin and Griffin Edwards, in an article in the Stanford Law Review, also used the Washington traffic-stop data to examine the impact of the resulting change in constitutional rights on racial bias in traffic stops. But they concluded that the data indicate that doctrines like Whren’s “contribute to a statistically significant increase in racial profiling of minority drivers.” As previewed above, our findings suggest no such effect.

---

25 Rushin & Edwards, supra note 21, at 637.
26 Id. at 643.
When presenting our findings, we discuss several areas of departure between our analysis and that of Rushin and Edwards. Primarily, we build on their work to implement a novel exercise designed to confront a fundamental challenge facing this investigation: the need to separate the effect of the change in Washington’s constitutional treatment of pretextual stops from its contemporaneous legalization of recreational marijuana. For these purposes, we take advantage of the fact that Colorado legalized recreational marijuana within days of Washington but did not alter its treatment of pretextual stops, allowing us to estimate a triple-differences specification that effectively uses Colorado as a control in our analysis.

Second, our analyses build on Rushin and Edwards’s with respect to another key challenge facing this exercise: the fact that driver race is not recorded for 27% of the stops in the dataset, an amount greater than the proportion of stops with non-white driver race recorded. One of Rushin and Edwards’s key approaches assumes that all drivers with missing race codes are non-white. We set forth two empirical markers that question the validity of this assumption. However, we find their alternative approach—dropping stops with missing race—less problematic. We nonetheless expand on that approach in several ways; primarily, we set forth imputation and other exercises designed to diagnose the extent of any possible bias that might arise from the missing-race data. These exercises suggest no concerns over any such bias.

Third—and perhaps speaking more specifically to why our findings do not align with those of Rushin and Edwards—we depart from their analysis in certain functional-form assumptions that we make in constructing our respective difference-in-difference and triple-differences analyses. Our reading of their approach is that they test for changes in absolute counts of traffic stops for non-white drivers before and after the Washington law change and compare those changes with observed changes in absolute counts of stops for white drivers before and after the reform. Given substantial baseline differences in population counts across races and thus in stop counts across races—especially in large counties—we elect instead to focus on changes in log counts of traffic stops. This allows us to make comparisons across races in the proportional (not absolute) responses in traffic stops to the Washington law change. This provides a more flexible approach that facilitates a comparison that is not sensitive to the scale of the outcome variable (and thus the baseline differences) and that arguably captures the

27 Our calculation differs slightly from Rushin and Edwards’s. See id. at 668 n.177 (“Around 26% of stops by the Washington State Patrol fail to list the race of the driver.”).
28 See id. at 669.
29 See id.
relationship of interest. We demonstrate that Rushin and Edwards’s results are rather specific to the functional form they select and are sensitive to this and related modifications.

Our analysis proceeds as follows. In Part I, we first lay out the legal backdrop for our inquiry. Then, we present and question the dominant position in the literature. We argue that a Fourth Amendment-type right against pretextual stops would do little to reduce racial disparities in traffic stops. In Part II, we introduce and analyze the Washington legal change that serves as the basis for our natural experiment. In Parts III and IV, we explain our natural experiment and results. In Part V, we discuss the implications of our results. First, we suggest that the exclusionary rule is an ineffective deterrent in the context of traffic stops. And second, we very briefly sketch out avenues for reducing racial bias in traffic stops that are worth exploring in the alternative.

I. THE QUESTION: WOULD A FOURTH AMENDMENT RIGHT AGAINST PRETEXTUAL STOPS REDUCE RACIAL DISPARITIES IN TRAFFIC STOPS?

A. Legal Background

Under the Fourth Amendment, police are generally able to pull someone over as long as they witness some traffic violation. The Fourth Amendment prohibits unreasonable searches and seizures by the government. And traffic stops are generally considered a Fourth Amendment seizure. A traffic stop is a permitted “reasonable” seizure if the officer has “‘reasonable suspicion’—that is, ‘a particularized and objective basis for suspecting the particular person stopped’ of breaking the law.” And one way people break the law is by committing traffic violations.

Until the Whren decision in 1996, there was some confusion in the federal courts about whether a pretextual stop was a reasonable seizure under the Fourth Amendment. Reasonableness is the “ultimate touchstone” of the Fourth Amendment. And one could argue that a stop, although identical to another stop that would be considered reasonable, is unreasonable when done for impermissible reasons. In other words, a stop is unreasonable when the proffered permissible reasons for the stop are merely pretextual. Alternatively, one could argue that as long as a stop could be justified by a

30 U.S. CONST. amend. IV.
32 Id. (quoting Prado Navarette v. California, 572 U.S. 393, 396 (2014)).
33 Lawton, supra note 22, at 922–23 (describing the circuit split). While there was a circuit split, the dominant position was that the Fourth Amendment did not prohibit pretextual stops. See id. at 922.
34 Heien, 574 U.S. at 60 (quoting Riley v. California, 573 U.S. 373, 381 (2014)).
permissible reason—e.g., the officer observed a traffic violation—the stop is reasonable, even if that reasonable basis was not the officer’s actual motivation. In *Whren*, the Court adopted the latter position.

The conflict in *Whren* began on a June evening in 1993.35 Plainclothes vice-squad officers in an unmarked car were patrolling a “high drug area” of D.C. Their “suspicions were aroused” when they saw two young Black men—James Brown and Michael Whren—in a truck with temporary plates stopped at a stop sign for more than twenty seconds. The officers noticed the truck’s driver looking towards the passenger’s lap. The officers made a U-turn to get closer. Brown and Whren turned right without signaling and drove away “at an ‘unreasonable’ speed.” When the police caught up to the truck, Officer Soto approached the driver’s window and saw two bags of cocaine in Whren’s hands. The officers arrested Brown and Whren and found more drugs in the car.

Brown and Whren sought to suppress the use of the drugs as evidence at trial.36 They did so by challenging the legality of the traffic stop. They argued that the officers did not stop them because they sped or because they failed to signal. Instead, they suggested, the officers actually stopped them because of an unsubstantiated hunch that Brown and Whren were up to some other illegal activity. In other words, they contended that their stop was pretextual and that pretextual stops are not allowed under the Fourth Amendment. Brown and Whren also argued that because “total compliance with traffic and safety rules is nearly impossible, a police officer will almost invariably be able to catch any given motorist in a technical violation.”37 So allowing pretextual stops would allow officers to “decide which motorists to stop based on decidedly impermissible factors, such as the race of the car’s occupants.”38 Thus, they suggested that the Fourth Amendment test should ask “whether a police officer, acting reasonably, would have made the stop for the reason given.”39

The Court held that pretextual stops do not violate the Fourth Amendment40—as long as there was some traffic violation, the stop is fine, even if it was really for some other reason, such as a hunch about criminal activity based on the driver’s race. The decision was unanimous.41 The Court

---

36 See id. at 809.
37 Id. at 810.
38 Id.
39 Id. (emphasis added).
40 Id. at 819.
41 Id. at 807.
characterized Brown and Whren’s “would have” test as too subjective and stated: “Subjective intentions play no role in ordinary, probable-cause Fourth Amendment analysis.” Thus, the Court concluded that intentional selective enforcement of the traffic code based on race was prohibited only under the Equal Protection Clause, not the Fourth Amendment. So instead of a “would have” test that would ask whether a reasonable officer would have made the stop in those circumstances, the Court endorsed a “could have” test, which simply asks whether there was a legal basis for the stop.

B. The Dominant Answer

Whren was met with a firestorm of criticism. One major critique has been that the decision contributes to racial bias in police stops and that a different holding—one that recognizes a Fourth Amendment right against pretextual stops—would reduce that racial bias. As Professor Margaret Lawton, reviewing the vast scholarly criticism of Whren, notes, “[M]any legal scholars have argued that use of a ‘reasonable officer’ test, also known as the ‘would have’ test, would combat the tendency to use traffic infractions as a pretext . . . and would consequently protect drivers.” For example, a recent empirical study demonstrating racial bias in traffic stops argued: “The Whren decision opened the floodgates to pretextual stops. Thus, tens of thousands of black and brown drivers have routinely been stopped and searched in an effort to reduce drug use.”

The basic logic of the argument of this dominant story is that when officers can engage in stops for reasons other than the traffic violation itself, stops are more likely to be racially biased. This outcome may arise from officers facing less restraint in acting on baseline racial biases, where such biases may either arise from racial animus or statistical discrimination along racial lines in pursuing non-traffic-related crimes. And as a result, per this

42 Id. at 813.
43 Id. Most agree that the Equal Protection Clause is an ineffectual tool for challenging a racially biased police stop. For one, “equal protection claims also are notoriously difficult to prove in criminal cases because they require a defendant to prove intentional discrimination on the basis of race between otherwise similarly situated individuals.” Brooks Holland, Race and Ambivalent Criminal Procedure Remedies, 47 GONZ. L. REV. 341, 346 (2011–2012); see Alison Siegler & William Admussen, Discovering Racial Discrimination by the Police, 115 NW. U. L. REV. 987, 991–93 (2021). And, importantly, the Equal Protection Clause does not provide a basis for excluding the fruits of an unlawful stop or impose some other penalty. See Holland, supra, at 346–47.
44 See, e.g., Baumgartner et al., supra note 21, at 108 (criticizing the Whren decision and suggesting that it has “opened the floodgates” to racially biased traffic stops); Chin & Vernon, supra note 21, at 941 (calling Whren “the Plessy of its era” and claiming that it “encouraged [the] spread” of racial discrimination by endorsing it); supra notes 21–22 and accompanying text.
45 Lawton, supra note 22, at 931.
46 Baumgartner et al., supra note 21, at 108.
dominant story, a Fourth Amendment rule that prohibits pretextual stops might reduce pretextual stops and thus reduce racial disparities in stops. We are not aware of scholarship affirmatively offering a counternarrative to this reaction to *Whren*.47

C. Reasons to Doubt the Dominant Answer

But we see several reasons to doubt the dominant account. We agree that pretextual stops will tend to fall unevenly across racial lines. However, even if the courts were to retreat from *Whren* and strengthen a Fourth Amendment right against pretextual stops, racial disparities in stops may not actually subside. Why? Because certain inherent features of this Fourth Amendment tool may simply fail to shape police behavior—that is, fail to deter officers from conducting pretextual stops. “A right is as big, precisely, as what the courts will do.”48 And in the context of traffic stops, the exclusionary rule—the Fourth Amendment’s primary enforcement mechanism—does little.

In this Section, we will first introduce and explain the two main elements of our theory of why a Fourth Amendment right against pretextual stops might not actually reduce bad stops—the difficulty in proving pretext and the fact that the remedy associated with a Fourth Amendment violation is not well suited to regulating an officer’s decision to initiate a stop. Then, to illustrate our theory, we will walk through the types of stops that are most likely to feature some racial bias and discuss the officer’s incentives in each case. At the outset, we emphasize that the reasons we provide to doubt the dominant account described above—and the resulting ambiguity in our expectations about the effects of a stricter Fourth Amendment standard on racial bias in stops—are precisely why we need careful empirical analysis. That is what animates this Article.

1. Proving a Violation

Even with a Fourth Amendment right against pretextual stops, defendants would likely find it difficult to prove that a given stop was pretextual. Often, the only witnesses to the stop will be the driver and the

47 That being said, the force of our Article does not entirely rest on the contention that those raising concerns of *Whren*’s possible contributions to bias are indeed the “dominant” view. And the main thrust of our empirical exercise is not to offer a formal, systematic review of this literature. Rather, we emphasize that based on our understanding of the literature, there is an important body of scholarship that has raised this view of *Whren*, and we are setting out to challenge this contention through the consideration of a quasi-experimental design. At an even more specific level, we are setting out to challenge the empirical support for this view of *Whren* that was set forth by Rushin & Edwards, supra note 21.

officer. In such cases, it seems unlikely that a court will conclude that the stop was pretextual. First, it is not inconceivable that officers might shape their testimony to fit the law, fudge facts, or even lie. Indeed, in a relatively recent series of interviews, judges and prosecutors in Chicago “willingly admitted that police perjury was part of the culture of the court system in Cook County.”

Second, judges might be inclined to take an officer’s word over the defendant’s. There is a long history of judicial deference to the police. For one, the judge’s view might be colored by the fact that the officer did in fact find something when searching the defendant. In a suppression hearing at which the Fourth Amendment objection would arise, the judge already knows that incriminating evidence was found, perhaps leaving the impression of an untrustworthy defendant. In addition, judges could have strategic reasons for siding with the police. Judges might want to avoid appearing weak on crime and maintain institutional relationships.

Building on the above points, of course, are the inherent difficulties in establishing subjective intent and demonstrating pretext. As we will discuss in Part II, Washington’s temporary retreat from Whren involved a rule with a subjective component, and our review of Washington case law revealed the difficulty in proving impermissible motive under a rule that required an interrogation of the officer’s actual motives. In particular, it seemed that


51 See Guido Calabresi, The Exclusionary Rule, 26 HARV. J.L. & PUB. POL’Y 111, 112–13 (2003) (suggesting that, in cases involving murderers and rapists, judges close their eyes and side with the police even when they suspect police perjury).


53 Christopher Slobogin, Why Liberals Should Chuck the Exclusionary Rule, 1999 U. ILL. L. REV. 363, 376; William J. Stuntz, Warrants and Fourth Amendment Remedies, 77 VA. L. REV. 881, 912 (1991) (“It must be much harder for a judge to decide that an officer had something less than probable cause to believe cocaine was in the trunk of a defendant’s car when the cocaine was in fact there.”); Nancy Leong, Making Rights, 92 B.U. L. REV. 405, 434–37 (2012).


absent an actual admission of an impermissible motive by the officer, it was quite uncommon for courts to find pretext, especially given the general deference afforded the police by the courts.56

Altogether, the deck seems stacked against defendants, even if a Fourth Amendment right was violated in truth. Even if the courts were to take a harsher Fourth Amendment stance against pretextual stops, these inherent evidentiary obstacles would weaken the bite of the Fourth Amendment as a potential tool and thus its ability to truly discourage officers from engaging in pretextual stops.

2. The Lack of a Meaningful Remedy

The exclusionary rule is the main remedy for a Fourth Amendment violation,57 especially in the context of traffic stops.58 The exclusionary rule says that any evidence acquired as a fruit of a Fourth Amendment violation cannot be used against the defendant at trial.59 The exclusionary rule was initially understood to be a constitutionally mandated remedy.60 But it has

56 See, e.g., State v. Cloe, No. 29007-3-II, 2003 WL 22137290, at *1–5 (Wash. Ct. App. Sept. 16, 2003). In that case, the officer testified that he stopped the defendant because he saw the defendant’s car turning off of 109th Avenue without using an indicator. In the course of the stop, the officer discovered bags of methamphetamine. The defendant, however, testified that he was never on 109th Avenue, instead claiming that he left a friend’s house on 110th Avenue and was stopped as he turned onto 112th Avenue, and he argued that the stop was pretextual. Even though the defendant set forth facts that undermined the basis for the alleged traffic violation and even though the officer’s testimony revealed that he knew that the friend’s house was a suspected drug house and that the area was a high-crime area, the trial court still found that the officer’s testimony was more credible than the defendant’s, and the appellate court dismissed the defendant’s pretextual-stop argument.

Our review of the case law did reveal some situations in which it appears that pretext was established, but only where that determination was supported by an officer admission (further reinforcing the point that an admission to this effect may be what it takes to prove pretext in a subjective-intent regime). See, e.g., State v. Cramer, No. 17953-2-III, 2000 WL 1663641, at *3 (Wash. Ct. App. Nov. 2, 2000). Interestingly, this case law review also suggested that how forthcoming officers may be in their testimony is sensitive to the prevailing legal regime. Most of these cases with officer admissions were cases in which the suppression hearing took place before the Ladson decision, but the appeal took place after, meaning that the admissions were made before the officer could have known that such an admission would affect the stop’s validity. See, e.g., State v. Moore, No. 43692-9-I, 1999 WL 1138575, at *2 (Wash. Ct. App. Dec. 13, 1999) (noting that the suppression hearing took place before the Ladson decision); see also infra text accompanying note 82 (discussing the Ladson holding). But once Ladson fully set in, we no longer saw such testimony—admittedly, our records review only looked at appellate opinions in which such testimony was discussed.

57 See Rachel A. Harmon, Promoting Civil Rights Through Proactive Policing Reform, 62 STAN. L. REV. 1, 10 (2009) (explaining that the exclusionary rule “is by far the most commonly used means of discouraging police misconduct and perhaps the most successful”).

58 Other remedies such as civil suits against the officer are ineffective in this context because the monetary compensation for an unlawful traffic stop is unlikely to be worth the effort and cost of a lawsuit for the average driver.


60 See id.
evolved into a remedy justified solely by its ability to deter police misconduct.\textsuperscript{61}

There are reasons to doubt, however, whether the exclusionary rule can shape police decisions about traffic stops, even if attached to a right against pretextual stops.\textsuperscript{62} First, for a punishment to shape behavior, it should regularly track the bad behavior. That is, when bad stops occur the punishment should regularly be imposed, and when good stops occur the punishment should not apply.\textsuperscript{63} But this does not happen in the context of traffic stops and the exclusionary-rule “punishment.”

Most traffic stops—good or bad—do not turn up incriminating evidence.\textsuperscript{64} Indeed, police might sometimes stop a driver with the intent to harass\textsuperscript{65} or simply to meet a quota.\textsuperscript{66} But the exclusionary rule only kicks in to exclude evidence that the state would like to introduce at a trial. So, finding incriminating evidence is a prerequisite to its application. Thus, even if the Fourth Amendment did prohibit pretextual stops, there is a good chance that violations might not be subject to even the possibility that the exclusionary rule will apply.

And even within the set of stops that do turn up incriminating evidence, there are several other reasons the exclusionary rule may not come into play. First, few resulting criminal cases will actually go to trial, where the potentially excludable evidence would be used. For one, in many cases the

\begin{itemize}
  \item [\textsuperscript{61}] See Davis v. United States, 564 U.S. 229, 236–37 (2011) (“The rule’s sole purpose, we have repeatedly held, is to deter future Fourth Amendment violations.”).
  \item [\textsuperscript{62}] Our argument is not that the exclusionary rule should not apply to traffic-stop-based evidence. Our point is much narrower: that because the exclusionary rule is weak in this area, Fourth Amendment-based attempts to regulate police traffic-stop behavior are unlikely to succeed. Many of the more general forms of our arguments here were first made in Slobogin, supra note 53.
  \item [\textsuperscript{63}] See id. at 373–74.
  \item [\textsuperscript{64}] See Richard M. Re, The Due Process Exclusionary Rule, 127 Harv. L. Rev. 1885, 1895 (2014); Harmon, supra note 57, at 10–11 (“[T]he scope of the exclusionary rule is inevitably much narrower than the scope of illegal police misconduct. The exclusionary rule provides a remedy only when police seek to use evidence that results from misconduct at a criminal trial. It therefore discourages officer misconduct only when the misconduct may produce evidence and when the government would value using that evidence at trial. Many kinds of misconduct do not have these characteristics. For example, Terry stops might be done primarily to harass or intimidate, and police uses of excessive force rarely produce evidence of a crime.”); Pierson et al., supra note 17, at 739.
  \item [\textsuperscript{65}] See Harmon, supra note 57, at 11.
\end{itemize}
police might not make an arrest.\textsuperscript{67} And even if an arrest is made, the case might not make it to trial. For instance, the defendant might take a plea deal.\textsuperscript{68} And even if the case does go to trial, the Fourth Amendment violation might not matter because it might not be raised. This is a distinct possibility in the context of pretextual stops because the defendant might not even suspect that the stop was not for the traffic violation.

And even if a Fourth Amendment objection is raised in a regime that prohibits pretextual stops, there might still fail to be a serious “punishment” incurred for the unconstitutional act. For one, the fruits of an unconstitutional search can still be admitted via exceptions to the exclusionary rule, such as the independent source,\textsuperscript{69} inevitable discovery,\textsuperscript{70} attenuation,\textsuperscript{71} and good faith\textsuperscript{72} carve-outs. Further, even if evidence is excluded, the jury might still find the defendant guilty on the weight of other nonexcluded evidence. And finally, even if the exclusionary rule kicks in, the evidence is excluded, and that defendant is found not guilty as a result, Fourth Amendment standing rules allow evidence obtained in violation of an individual’s Fourth Amendment rights to be used against someone else in court.\textsuperscript{73} So there is a possibility that the fruits of a judicially recognized unconstitutional stop can still be used to get a conviction against someone.

The upshot is that the exclusionary rule might not be an effective punishment tool in the vast majority of bad stops. And with the threat of exclusion in the context of traffic stops not being a serious ex ante threat, it is unlikely to meaningfully shape police behavior.\textsuperscript{74}

The exclusionary rule’s inadequacies in the context of traffic stops do not stop there. Even if a punishment properly tracks the desirable and undesirable behavior, for it to shape behavior and deter the targeted act, it has to be seen by the actor as negatively impacting that actor. There are numerous reasons for skepticism here.

When deciding whether to initiate a traffic stop, it is possible that many officers do not care much about whether their stop results in a conviction.

\textsuperscript{67} See Slobogin, supra note 53, at 374–75; see also Terry v. Ohio, 392 U.S. 1, 14 (1968) (stating that the exclusionary rule “is powerless to deter invasions of constitutionally guaranteed rights where the police either have no interest in prosecuting or are willing to forgo successful prosecution in the interest of serving some other goal”).

\textsuperscript{68} But a defendant’s sense about whether evidence will be excluded might shape his calculation about whether to take a plea deal and what sort of deal to take.

\textsuperscript{69} Murray v. United States, 487 U.S. 533, 537 (1988).

\textsuperscript{70} Nix v. Williams, 467 U.S. 431, 444 (1984).


\textsuperscript{72} Herring v. United States, 555 U.S. 135, 142 (2009).


\textsuperscript{74} See Slobogin, supra note 53, at 373–80.
Often, officers care much more about arrests than convictions—a sense that is reinforced by the metrics that determine career advancement. In the context of traffic stops, the officer might care more about getting a certain number of stops on the books to meet a quota than about whether the stop will result in finding and prosecuting criminal activity. Further, many officers feel that they lack control over whether their arrest results in a conviction because they associate the failure to convict with a litany of external factors. As a result, some officers might not see a nonconviction as a punishment for their Fourth Amendment violation. Instead, driving their decision over whether to initiate a stop are various other factors.

Moreover, some officers may care about a conviction but may see little cost arising from the exclusion of evidence resulting from a pretextual stop. After all, even if some evidence that is found gets excluded, the officer might never have gotten that evidence absent the stop in the first place. Accordingly, they may see that outcome as not actually negatively impacting the overall chances of conviction.

In sum, given the difficulties involved in proving pretext and given the weaknesses of the exclusionary rule as a remedy that is likely to influence officers, it is questionable whether the Fourth Amendment can serve as a tool that effectively shapes officers’ traffic-stop behavior. To the extent this is true, it is therefore questionable that Whren can be seen as meaningfully contributing to the racial disparities in traffic stops that unquestionably exist.

3. Our Theory in Context

To further explain and illustrate our theory of why a Fourth Amendment right against pretextual stops will not actually reduce bad stops, we conceptualize several different types of stops that are likely to contribute to racial disparities and then assess the potential for the Fourth Amendment to deter racial bias in each scenario. In doing so we demonstrate that there are very few hypothetical instances in which a Fourth Amendment right against pretextual stops is likely to change police behavior.

The first type of traffic stop that we conceptualize as contributing to racial disparities is when the officer intends to be unbiased in deciding whom to stop and intends to enforce traffic violations specifically, but because of unconscious biases the officer disproportionately stops non-white drivers. In these cases, the officer has no reason to consider whether her stop is legally

---

75 See id. at 377–78.
76 See id. at 378.
77 Cf. id. at 377–78 (“But the sociological literature strongly suggests that the primary goal of officers in the field in the average case is to get a ‘collar.’ If they do, they’ve done their job. It is the prosecutor’s job to convict.” (footnote omitted)).
78 See id. at 378.
problematic because she herself does not know that she is stopping the driver for something other than the traffic violation and does not know that she is doing so differentially based on race. And because she is primarily motivated by a desire to enforce the traffic code in a uniform manner, she is unlikely to consider whether evidence she might uncover in the course of the stop will be inadmissible in a future trial. Thus, the exclusionary threat is unlikely to shape her decision about whom to stop.

The second hypothetical scenario is when the officer intends to be biased in her stops simply to harass non-white drivers. Take for example the candid statements of a state trooper regarding a stop that happened where we both live—Durham, North Carolina. When the trooper was asked why he stopped the driver for a seat-belt violation, he responded:

“Everyone knows that a Hispanic male buying liquor on a Friday or Saturday night is probably already drunk”; “Mexicans drink a lot because they grew up where the water isn’t good”; and that he did not care what happened in court “as long as I get them [(i.e. Hispanic males)] off the road and in jail for one night.”

When asked whether he targets Hispanics, the trooper said: “I’m not targeting Hispanics. Most of my tickets go to blacks.” When officers’ goals are primarily to harass, stopping and harassing may be good enough for them. The possibility of the exclusionary rule barring evidence from the stop at trial may simply be an inconsequential input into their decision-making process.

The third hypothetical type of race-based stop we consider is when the officer needs to pull some people over for traffic violations—e.g., to fulfill a quota—and determines that she might as well pull over drivers that she deems are most likely to have contraband or other evidence of crime. Either consciously or subconsciously, this hypothetical officer may then feel that she is most likely to find such evidence on non-white drivers. The application of a harsher “would have” Fourth Amendment test may have a greater chance of deterring this officer from proceeding with the racially biased pretextual stop relative to the hypothetical officers from the first two scenarios. However, even in this scenario, it is perhaps unlikely that a Fourth Amendment right against pretextual stops will do much for the various reasons set forth in the preceding Section: (1) the officer may care about the possibility of exclusion, but she might severely discount its probability given the difficulties in proving pretext; (2) the officer may not feel that there is a

80 Id.
real cost to the exclusionary rule since, even if the evidence is excluded, it might never have been retrieved absent the stop; or (3) the officer may not care about the possibility of exclusion if her motivations are focused more on arrests than convictions.

The fourth type of race-based stop that we consider arises when a hypothetical officer only cares about criminal activity other than traffic stops and where the officer initiates a stop based on a hunch that is tainted by racial bias—e.g., they may pull over a non-white driver based on their assessment that the crime in which they are interested is more likely to have been committed by a non-white actor. Out of the four scenarios discussed, this is the one in which it is most likely that the application of a harsher “would have” Fourth Amendment test would deter the hypothetical officer from executing the pretextual stop.\textsuperscript{81} Unlike the first scenario discussed, this hypothetical officer is well aware that her actions might trigger legal scrutiny, as her hypothesized motivation is indeed to engage in pretextual behavior. Moreover, relative to the third scenario just discussed, it is perhaps more likely that this hypothetical officer will care about the exclusion of evidence and about the conviction of the suspect given our assumption regarding her primary objectives to pursue the non-traffic-related crimes.

However, for some of the other reasons just raised in the third scenario, there are still reasons to doubt whether even the hypothetical officer in this fourth scenario may be deterred by a “would have” Fourth Amendment test that prohibits pretextual stops and that enforces this prohibition via the exclusionary rule. For one, the officer may still discount the probability of the exclusionary rule being applied given evidentiary barriers that are still likely to apply in this instance—e.g., if the officer does not admit that the reason for the stop was pretextual, it might be very difficult to prove. Moreover, as above, even if the officer thought that there was a real risk that a judge would say the stop violated the Fourth Amendment, the officer may not deem the exclusion of any evidence a significant cost. For DUI and drug-possession hunches, the driver will probably be long gone before the officer can develop the requisite suspicion to justify the stop. Accordingly, even if the fruits of the stop are inadmissible, she would not have been able to get the evidence without the stop and is thus in no worse of a position.

\textsuperscript{81} As a starting point, we note that the application of a “would have” test could lead to the exclusion of any evidence arising from a stop under the scenario. Stepping back, under \textit{Whren}'s “could have” test, as long as there was some traffic violation, these stops are upheld. \textit{See supra} notes 41–43 and accompanying text. But, under a “would have” test, a defendant could argue that an officer who was not acting based on an unsubstantiated hunch would not have pulled him over. \textit{See supra} note 39 and accompanying text.
All of this is of course highly speculative. We do not know how many stops are for each of these reasons or whether there are other significant types of racially biased stops that escaped us. Nor do we know for a fact that we have accurately captured officers’ decision-making processes in these situations—which are unlikely to be monolithic. Our main objective with this discussion is to simply raise doubts regarding the conventional wisdom that the more permissive “could have” approach embraced by Whren is contributing to racial disparities in traffic stops and that a harsher “would have” approach would instead reduce such disparities. We now endeavor to help resolve this ambiguity by turning to a natural experiment born out of a change in constitutional law in the State of Washington.

II. WASHINGTON AS A NATURAL EXPERIMENT

Three years after the U.S. Supreme Court in Whren held that the federal Constitution’s Fourth Amendment allows pretextual stops, the Washington Supreme Court held that Washington’s constitutional analogue to the Fourth Amendment prohibits pretextual stops.82

That case—State v. Ladson83—sets the stage for the natural experiment we investigate in this Article. Ladson reasoned that the Washington constitution generally bars warrantless searches and seizures. One exception is for traffic stops based on reasonable suspicion that a traffic violation has occurred. The reason for that exception is based on the premise that such stops are done to enforce the traffic code—a sort of community caretaking exception to the general prohibition on warrantless seizures. The exception, the court stated, was not made so that police could better investigate criminal activity. Indeed, in Washington, traffic violations had been decriminalized. Thus, a stop in response to a traffic violation is by definition not a stop to investigate criminal activity. In fact, the Washington Supreme Court noted that part of the reason the state legislature decriminalized traffic violations was to prevent pretextual stops. So the court was unwilling to allow a community caretaking exception to the warrant requirement to serve as a backdoor to an investigative stop. As a result, the court held that the Washington constitution’s analogue to the Fourth Amendment prohibits pretextual stops.84 It suggested that a stop is pretextual when the traffic violation (or the constitutionally permissible reason for the stop) is not “the

82 State v. Ladson, 979 P.2d 833, 842 (Wash. 1999) (en banc).
83 See id. at 837–39, 841–42 (recounting the legal background of the case).
84 Id. at 842.
true reason” for the stop.\textsuperscript{85} To determine whether a stop is pretextual, the \textit{Ladson} court instructed that a reviewing court should consider both the objective reasonableness of the stop and the officer’s subjective intent.\textsuperscript{86}

Thirteen years after \textit{Ladson}, in December 2012, the Washington Supreme Court pulled back from its broad language and instated a standard functionally equivalent to \textit{Whren}. In \textit{State v. Arreola}, the Washington Supreme Court purported to maintain that pretextual stops are unconstitutional but severely narrowed what counts as a pretextual stop. \textit{Arreola} held that a mixed-motive stop—a stop done for both permissible and impermissible reasons—is not a pretextual stop.\textsuperscript{87} As long as the traffic infraction was “an actual, conscious, and independent cause of the traffic stop,” the stop does not violate Washington’s Fourth Amendment analogue.\textsuperscript{88} This is true even if the primary reason for the stop was an impermissible one\textsuperscript{89}—such as an unsubstantiated hunch based on the driver’s race.

For our analysis of the effect of the \textit{Arreola} decision on racial disparities in traffic stops to be informative on the effects that \textit{Whren} has had, it needs to be the case that the \textit{Ladson}-to-\textit{Arreola} shift represented a true legal change—i.e., that the Washington courts between \textit{Ladson} and \textit{Arreola} indeed applied a harsh standard against pretextual stops and that the courts post-\textit{Arreola} indeed relaxed the standard employed. To confirm this, we reviewed every Washington appellate case citing either \textit{Ladson} or \textit{Arreola}. They tell us that there was a real change in Washington’s constitutional law. They also tell us that the \textit{Ladson} test was stricter on police officers than either \textit{Whren} or the “would have” test, while the \textit{Arreola} test is the functional equivalent of \textit{Whren}. Let us elaborate.

Under \textit{Ladson}, the traffic violation had to be the actual reason for the stop. As an illustration, take \textit{State v. Capshaw}, a pre-\textit{Arreola} case. There, the appellate court agreed with the trial court that the relevant test under \textit{Ladson} is “whether the facially valid reason for the stop is ‘sufficiently primary’ among the various possible reasons for the stop.”\textsuperscript{90} In other words, to be legal, the traffic violation needed to have been more than just a reason for

\textsuperscript{85} See id. (“We conclude the citizens of Washington have held, and are entitled to hold, a constitutionally protected interest against warrantless traffic stops or seizures on a mere pretext to dispense with the warrant when the true reason for the seizure is not exempt from the warrant requirement. We therefore hold pretextual traffic stops violate article I, section 7, because they are seizures absent the ‘authority of law’ which a warrant would bring.” (citing WASH. CONST. art. I, § 7)).

\textsuperscript{86} Id. at 843.

\textsuperscript{87} See 290 P.3d 983, 991 (Wash. 2012).

\textsuperscript{88} Id.

\textsuperscript{89} Id. at 991–92.

the stop. Another example is *State v. Meckelson*. There, the court asked, “Was the officer’s stop *solely* for the driver’s failure to signal?”

This is different than the *Arreola* standard because under *Arreola* it does not matter how important the traffic-infraction motivation for the stop is compared to other motivations as long as the traffic-infraction motivation was *sufficient* to be an actual, conscious, and independent reason for the stop. The post-*Arreola* cases confirm this. Take, for example, *State v. Burr*, in which the court said, “Even if . . . [the impermissible reason] was the primary reason for the stop, under *Arreola*, this inference does not alter our analysis . . . [so long as the officer] had an actual, conscious, and independent reason to stop Burr based on the traffic infraction.”

Thus, the cases confirm the following descriptions of the tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Standard</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Whren</em></td>
<td>“Could have” test: Is there a legal basis for the stop?</td>
<td>Purely objective</td>
</tr>
<tr>
<td>“Would have”</td>
<td>Would a reasonable officer have made the stop in those circumstances?</td>
<td>Mostly objective</td>
</tr>
<tr>
<td>test</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Arreola</em></td>
<td>Was the legally permissible reason for the stop (the traffic violation) an actual, conscious, and independent reason for the stop?</td>
<td>Subjective plus objective</td>
</tr>
<tr>
<td><em>Ladson</em></td>
<td>Was the legally permissible reason for the stop (the traffic violation) the actual reason for the stop?</td>
<td>Subjective plus objective</td>
</tr>
</tbody>
</table>

*Ladson* is the strictest test. It easily takes a harsher stance on pretextual stops than either *Whren* or *Arreola* because under *Ladson*, even if the traffic violation was actually a reason for the stop, it might still be unconstitutional. *Ladson* even surpasses the “would have” test that scholars have endorsed. Under *Ladson*, even if a reasonable officer would have made the stop in those circumstances, if the officer did not in fact make the stop for the traffic violation, the stop is unconstitutional.

Meanwhile, *Arreola* is functionally equivalent to *Whren*. Under *Whren*, there need only have been a traffic violation. Under *Arreola*, there must have been a traffic violation and that violation must have at least crossed the officer’s mind when deciding whether to make a stop. But there is likely no meaningful difference between the two tests. Under either test, the officer has to make sure there is a traffic violation to make the stop. So the traffic violation will almost always be on the officer’s mind when she makes a stop.

---

Thus, the traffic violation will always be a reason for the stop, making Arreola’s additional subjective requirement largely superfluous. A functional difference between Whren and Arreola would probably only arise in cases in which an officer admits that the only thing on her mind when making the stop was the impermissible reason for the stop (though a reasonable officer could have executed the stop). In that case, if a traffic violation had in fact occurred, the stop would be upheld under Whren but struck down under Arreola. But such a situation seems exceedingly unlikely. Indeed, we found no cases post-Arreola in which a traffic stop was deemed unconstitutional under Washington’s analogue to the Fourth Amendment.

Because Arreola meaningfully altered the standard applied in adjudicating pretextual stops, the legal change in Washington presents a good opportunity to test whether a Fourth Amendment right is likely to affect racial bias in police stops.

For example, imagine a pre-Arreola Washington in which a state patrol officer received a tip of a drug purchase that had just occurred in a neighborhood. The officer noticed a car driving in the neighborhood that the officer had a hunch might be the car driven by the purchaser, but the officer did not think that her hunch amounted to reasonable suspicion that would justify the stop. Nonetheless, the officer stopped the driver, citing a broken taillight, a minor traffic infraction. If the reviewing court felt that the main reason for the stop was to investigate the drug purchase and not the traffic infraction, the stop would be deemed unconstitutional and any evidence about the drug purchase obtained during the stop would be thrown out under Ladson. But after Arreola, as long as the officer can say that one actual, conscious, and independent reason for the stop was the broken taillight, then the stop will be held constitutional. This is true even if the broken light was not the primary reason for the stop. Now, imagine that the officer’s hunch over the drug purchase was racially biased—e.g., the officer perceived that non-white drivers are more likely to purchase drugs in the neighborhood in question. In this case, the greater discretion extended by Arreola to enact a stop based on this hunch may contribute to greater racial disparities in the initiation of stops.

Accordingly, the Arreola decision—by capturing real variation in the stringency of Fourth Amendment protection—offers an opportunity to test the dominant theory that this protection may indeed have a bearing on racial disparities in stops. To the extent we do document such a responsiveness, that result would tend to support the dominant concern that the permissive “could have” approach embraced by Whren is causing a greater degree of racial disparities in stops than would otherwise be present under a more
restrictive standard. Of course, just because we have validated that courts will apply a lightened standard in the Arreola era does not mean that officer behavior will respond in the way predicted by the dominant theory. This was the exact point of Part I of this Article. No matter how stringent one sets the standard regarding pretextual steps, there may be certain inherent features of this Fourth Amendment tool that dampen the deterrent channel connecting the court proceedings to the officer’s decision-making process. Again, it is this doubt that will motivate our empirical analysis, to which we now turn.

III. DATA AND PREVIOUS EMPIRICAL LITERATURE ON ARREOLA

A. Data

Our empirical analysis of the effects of Arreola on racial bias in the initiation of traffic stops will rely upon traffic-stop data from several jurisdictions. Each of these jurisdictional databases was provided by the Stanford Open Policing Project. Given our intent to explore the effects of a change in law in Washington, our primary source of data on traffic stops is the Washington State Patrol database. As will be discussed below, we focus on analyzing stops occurring between 2010 and 2015 in our main analyses. These data record various pieces of information on the over 7 million stops occurring over that time, including (1) the date and time of day (in one-hour increments) of the stop, (2) the county of the stop (in addition to location information at the subcounty level), and (3) the driver’s race, sex, and age.

To implement certain secondary analyses—as we will discuss in greater depth below—we also draw on traffic-stop data from 2010 to 2015 from the Colorado State Patrol (roughly 2.5 million stops) and from the Seattle Police Department (roughly 160,000 stops). Data for these additional jurisdictions also come from the Stanford Open Policing Project. In certain robustness checks, we extend the above data sources until 2017 and 2018, as we discuss further below.

B. Prior Analysis and Preview of Findings

Before getting into our analysis, we note that we are not the first to analyze the Washington traffic-stop data before and after Arreola to investigate the potential for a right against pretextual stops to combat racial bias in police stops. Professors Rushin and Edwards analyzed the data first and conclude “Arreola . . . [is] associated with statistically significant

93 Data, STAN. OPEN POLICING PROJECT, https://openpolicing.stanford.edu/data/ [https://perma.cc/YS29-68L5]. The data were compiled in connection with Pierson et al., supra note 17.
94 Data, supra note 93.
increases in stops of Black drivers, Hispanic drivers, and drivers of ‘other’ races relative to white drivers.”95 They go on to suggest that if the Court had ruled differently in Whren, that might have “reduced racial bias by officers.”96 As Rushin and Edwards state, if their findings gain wide acceptance, Ladson’s holding might serve as a blueprint for regulating pretextual stops.97

By way of preview, our findings differ from those presented by Rushin and Edwards. Our own findings can be interpreted in one of two ways. In the most aggressive light, our findings suggest that Arreola made no marginal contribution—or at most a modest contribution—to racial disparities in officers’ decisions to initiate traffic stops, which would suggest that our solution to the undeniable problem of racial bias underlying such stops should not lie with a Fourth Amendment-like approach. In a more conservative light—considering the challenges involved in distinguishing a “null effects” story from a “zero effects” story—our results can be read as suggesting that, when taking a range of robust approaches to testing the effects of Arreola on racial disparities in stops, we fail to find evidence to suggest that it increased such disparities.98

From either perspective, our results cast substantial doubt on the conclusions reached by Professors Rushin and Edwards in their investigation into Arreola’s effects on traffic stops. In the analysis and results that we present below, we address the various ways in which our respective analyses depart from one another.

95 Rushin & Edwards, supra note 21, at 669.
96 Id.
97 Id.
98 One argument in favor of this more conservative takeaway is that the approach considered by Rushin & Edwards, supra note 21, and that we build upon still relies upon just one state that has modified its constitutional protections against pretextual stops. While that fact alone does not raise concerns over biased estimates, it increases concerns that spurious shocks to the stop rates of non-white drivers may happen to coincide with the post-Arreola period and confound our analysis. With a large number of treatment states, these spurious shocks would be more likely to average out to zero. See Timothy G. Conley & Christopher R. Taber, Inference with “Difference in Differences” with a Small Number of Policy Changes, 93 REV. ECON. & STAT. 113, 113 (2011) (noting that point estimates in difference-in-difference regressions with a small number of policy changes are statistically inconsistent). Further challenges arise with a small number of treatment states when estimating the standard errors associated with the difference-in-difference coefficients. See id. Of course, the two-alternative triple-differences specifications estimated in Part IV are designed to address some of these concerns, though those empirical models likewise require certain assumptions to interpret the results in causal terms, as we set forth below. For these reasons, we note that the most conservative reading of our results is not in establishing that Arreola had no effect on racial disparities in stop rates but that we have failed to find evidence of any such effects through the methods we have available.
IV. ANALYSIS

A. Overview of Empirical Investigation

To recap our discussion above, our aim in this Article is to test the dominant theory that a permissive constitutional standard regarding pretextual stops—such as that set forth in Whren—contributes to increased stops of non-white drivers. To test this theory, we draw upon the 2012 Arreola decision by the Washington Supreme Court, which took the State of Washington from a stringent regime that prohibited pretextual stops to a more permissive regime that is functionally equivalent to that of Whren. Accordingly, we aim to test in this Part whether stops of non-white drivers increased following Arreola. Our analysis will demonstrate that no matter how we address certain empirical challenges, we consistently fail to find evidence to suggest that Arreola contributed to an increase in the rate of stops of non-white drivers (or Black drivers specifically).

At the outset, we reiterate that we are assuming that racial bias in traffic stops does exist. That is, we do not believe that a reasonable interpretation of the null results that we present is that there is no underlying racial bias in the first place. Any such interpretation would simply be inconsistent with the world around us. Similarly, we do not set out in this Article to empirically demonstrate that there is an underlying racial basis in traffic stops—that is, we do not try to test a null hypothesis that such bias does not exist.99 Taking the existence of such bias as unquestionable, our empirical inquiry instead will be focused on assessing whether and how the law may be able to address this fundamental problem.

Before diving into this analysis, we caution that our analysis—though it attempts to draw broader inferences about the effects of Fourth Amendment-like treatments of pretextual stops—is based on the analysis of just one state intervention. We acknowledge that this leaves us with both internal and external validity concerns. For instance, regarding external validity, even if our analysis accurately identifies the effect of substantive constitutional treatment of pretextual stops on officer behavior in Washington, that effect may not generalize to officers in other states. While we concede this possibility, we note that we have no particular reason to doubt the generalizability of our findings.

99 For the same reason, we also do not set forth a robust literature review on those studies that have attempted to demonstrate racial bias in traffic stops. That said, we do make note of a recent study that has at least demonstrated that this bias is pervasive across officers, rather than just being concentrated among a few “bad apples.” See Felipe Goncalves & Steven Mello, A Few Bad Apples? Racial Bias in Policing, 111 AM. ECON. REV. 1406, 1426 (2021) (finding that 42% of officers are discriminatory in a study analyzing data from the Florida Highway Patrol on the charged speed for speeding tickets).
Regarding internal validity concerns stemming from our Washington-centric analysis, consider one of the key challenges we will face below—i.e., to distinguish between an effect of Arreola on racial disparities in traffic stops and the effects of other shocks that may hit Washington in the post-Arreola period that also impact the rate by which non-white drivers may be stopped. If we were able to observe more states experimenting with constitutional treatment of pretextual stops in the way Washington did, we would be able to build an even richer quasi-experiment to help separate the causal effect of the constitutional treatment from the effects of other factors that may impact the stop rate of non-white drivers. If our only approach were to compare the stop rates of non-white and white drivers within Washington before and after Arreola, we agree that our focus on just one state would be a limitation. Our analysis goes deeper than that, however, in ways that will allow us to address matters of this nature and better target the effects stemming from the Arreola decision itself. As we will discuss below, we will estimate two types of triple-differences models that will help us achieve this necessary separation.

B. Difference-in-Difference Analysis: Overview

As set forth in Part I, according to the dominant theory, Arreola’s permissive standard for pretextual stops would lead to an increase in the rate of stops of non-white drivers (or of Black drivers, specifically). If we were to simply test whether the rate of non-white stops increased after Arreola, we would immediately confront a key empirical challenge. Traffic-stop rates will change over time for other reasons. For instance, perhaps state-trooper rosters increased or decreased over this time frame. Perhaps people drive more or less frequently over time. If we had data on all such developments, we could, of course, simply include them as control variables in a regression analysis. However, even if we could collect data on some such stories, we may remain concerned over what we cannot observe.

To confront this immediate challenge, we will explore how the rates of stops of non-white drivers relative to that of white drivers change before and after Arreola (in the alternative, we will focus on Black drivers, specifically). That is, we will look at the difference between non-white and white stop rates and analyze how that difference evolved over time as Washington’s constitutional treatment of pretextual stops changed. In essence, this strategy uses white drivers as a control group—i.e., it uses the change in the rate of stops of white drivers after Arreola as a way to estimate those unobservable factors that are changing over time and that may impact traffic-stop rates. Having done so, we can then subtract that before–after white-driver effect from the before–after change for non-white drivers in order to isolate the
effect of *Arreola* and net out the influence of the confounding factors. An empirical strategy of this nature goes by the name of a “difference-in-difference” design.\(^{100}\)

Of course, for this strategy to rule out the influence of these unobservable factors, one must assume that such factors affect non-white and white drivers alike. For instance, one must assume that changes in state-patrol rosters over time will affect non-white stop rates the same way they affect white stop rates. In other words, this strategy will allow for trends in non-white stop rates and white stop rates that bounce up and down in parallel over time due to unobservable shocks that are common to both groups. It will then attempt to explore whether these trends stop moving in parallel after the *Arreola* decision and whether those trends instead begin to diverge as the gap between stop rates widens.

In the next Section, we will begin to lay out difference-in-difference findings of this nature. We note, however, that we will return to the assumption set forth in the preceding paragraph after presenting these preliminary results. We will then discuss strategies that we employ that will allow us to relax that assumption and to estimate the effect of *Arreola* while accounting for unobservable factors that change over time and that are even specific to given races.

Before laying out these preliminary findings, we make one additional preliminary note regarding the time frame of our investigation. In our difference-in-difference exercises, we focus on a three-year postreform period. We acknowledge that one limitation of this evaluation window is that it may preclude the observation of a true effect that happens to occur with a lag longer than three years—e.g., because officers may not begin to respond to the altered legal landscape for at least three years. This is a valid caveat, but we must balance it against one of the limitations of the difference-in-difference approach, especially when this approach draws on just one treatment state. Most specifically, our key challenge with this difference-in-difference design comes in assuming that the white and non-white stop rates would have trended in the same manner absent the *Arreola* decision. The longer out we look, the more dubious this assumption becomes given that the longer out we look, the more likely it is that we will encounter another reform or occurrence in Washington that will lead to differential effects in stop rates across race.\(^{101}\) To balance our concerns over possibly lagged


\(^{101}\) With a greater number of treatment states, we could perhaps more reliably estimate longer run effects, as spurious shocks to stop rates would likely average out to zero.
officer responses and Washington-specific reforms that have differential racial impacts on stop rates, we select a three-year window in our primary approach.

That said, this balance is arguably altered in our triple-differences designs—e.g., when we further compare rates in Washington and Colorado and when we compare rates during the day and overnight. These designs employ additional layers of control to address concerns about other Washington-specific reforms that differentially affect stop rates by race, arguably providing us with greater faith that a longer postevaluation window will reflect responses to Arreola and not responses from unobserved Washington-specific shocks. Accordingly, in this richer triple-differences specification, we also show results using a longer evaluation window.

C. Difference-in-Difference Analysis: Findings

1. Preliminary Analysis: Raw Plots of Stop Rates

Before formalizing this difference-in-difference strategy in a regression framework, we will begin by using the raw traffic-stop data and plotting simple time trends of non-white and white stop rates over time (and alternatively, the stop rates of Black and white drivers over time). In the spirit of the difference-in-difference approach just discussed, this will allow us to assess whether we indeed observe a divergence in these trend lines that appears to begin following the Arreola decision. We will plot these trends in monthly intervals from 2010 to 2015—that is, three years prior to and subsequent to the December 2012 Arreola decision—using state-trooper data from the entire state.

Simply plotting trends in the total number of stops of non-white and white drivers may be misleading to the extent non-white and white population counts differ. More useful will be trends that reflect the number of stops that a given driver of each racial group is likely to experience. Accordingly, we plot trends in the rate of traffic stops, normalizing the stop count by the relevant population—e.g., the number of white drivers per white population. For these purposes, we normalize by adult population counts,\(^\text{102}\) as this approximates the driving population (though, all of the results that we present are robust to simply using the total population across all ages).

\(^{102}\) We acknowledge that one concern with normalizing by adult population is that the adult population does not necessarily perfectly capture the at-risk population of more fundamental interest—e.g., the number of people who drive by race. Nonetheless, including no normalization at all does not strike us as a more reasonable approach than normalizing by adult population. The Online Appendix, in any event, demonstrates the stability of our findings to these choices. See Figure A7 and accompanying text, available at https://scholarlycommons.law.northwestern.edu/nulr/vol116/iss6/2/. For instance, we also present results with no normalization, which reinforce the conclusions that we make in this Article.
**Figure 1: Concurrent Trends in Non-White- and White-Traffic-Stop Rates, for Three Years Prior to and Subsequent to *Arreola* Decision in Washington**

![Graph showing concurrent trends in non-white and white traffic stop rates.](image)

**Figure 2: Concurrent Trends in Black- and White-Traffic-Stop Rates, for Three Years Prior to and Subsequent to *Arreola* Decision in Washington**

![Graph showing concurrent trends in black and white traffic stop rates.](image)
We begin in Figure 1 by plotting trends over time in non-white and white stop rates in the same graph, with the stop rate indicated on the y-axis and the time period—characterized by months prior to and subsequent to the Arreola decision—on the x-axis. In Figure 2, we do the same but plot concurrent trends in the stop rates of Black drivers and white drivers.

Several observations emerge from these Figures. First, the rate of stops of Black drivers per population exceeds that of white drivers, though the stop rate of white drivers exceeds that of non-white drivers as a group. Second, and most critical for this Article, it does not appear that the differential in the non-white (or Black) rate and the white rate appears to change meaningfully as a result of the Arreola decision. If anything, Figure 1 appears to demonstrate a slight convergence in these trend lines after Arreola, in contrast with the divergence predicted by the dominant theory that a Whren-like approach to pretextual stops contributes to racial bias. If we take a closer look at Figure 2, it does suggest a subtle divergence in Black and white stop rates over time. However, this divergence appears to predate Arreola, casting doubt on whether it could have been caused by Arreola.

While the Arreola decision substantively changed the legal standard applied by courts, this preliminary observation of trends in stops by race around this decision lends support to the countertheories set forth in Part I. That is, the difficulty in proving pretext and weakness of the deterrent effect of the exclusionary rule may indeed prevent these substantive developments in the law from altering officers’ behavior. In the analysis to follow, we subject this analysis and conclusion to far greater scrutiny. In short, we will explore whether (1) Arreola truly did not alter the preexisting pattern of racial bias in stops, as Figures 1 and 2 suggest, or (2) Arreola in fact increased racial disparities in stop rates but other variables missing from this preliminary analysis also affect racial disparities in stops and are masking—i.e., working in the opposing direction—the true impacts of the change in the law stemming from Arreola.

2. **Difference-in-Difference: Overview of Regression Implementation**

To assess whether other factors may indeed confound the conclusions implied by Figures 1 and 2, we now turn to formalizing the difference-in-difference strategy discussed in Section IV.B via regression analysis. There are three key benefits of taking a regression approach over simply eyeballing Figures 1 and 2 and assessing whether we begin to see divergence in the non-white and white trend lines following Arreola.

First, the parameters of interest that we will estimate in the regression will directly produce a time trend in the differential non-white–white stop rate. This will obviate the need to visualize and discern the difference between the two separate lines from Figures 1 and 2 and then to see how that
difference evolves. Rather, we can produce a single trend line—i.e., the trend in the differential stop rate—and can simply observe if that trend line begins to move up after *Arreola*. Second, by turning to a regression approach, we are provided with a straightforward way to produce confidence intervals for the differential in the non-white–white stop rates. This will facilitate our ability to make statistical inferences.

Third and most importantly, by using multivariate regression, we are able to build on our observation of Figures 1 and 2 and account for the influence of other factors that (1) we can observe in the data, (2) may also be changing over time, and (3) may also affect the differential in non-white and white stop rates (or the differential in Black and white stop rates, specifically). To be clear, the difference-in-difference structure already flexibly accounts for factors that change over time that are common to both Black and white drivers across the state, as we discussed above. What we are referring to in this paragraph is the additional ability to directly account for the influence of factors that we can specifically observe in the data and that may happen to differentially affect stop rates across races.

For instance, one of the key controls that we will include in our regression is the county in which the stop occurs. Figures 1 and 2 convey that *Arreola* did not seem to lead to greater divergence in the stop rates of non-white and white drivers across Washington. However, hypothetically, it could be true that *Arreola* caused police officers to increase their propensity to stop non-white relative to white drivers. Indeed, it could also be true that around the time of the *Arreola* decision, the Black population in Washington began to reside more in counties that tended to have fewer stops of drivers overall—Black and white alike. If you combine these hypothetical developments, you might see that on net, there is no change in the rate by which Black drivers are stopped relative to white drivers even though in this hypothetical, *Arreola* did have an effect that changed officer behavior and intensified racial bias on any given stop. This hypothetical highlights the benefit of controlling for the county of the stop in the regression.

Papers employing difference-in-difference techniques will often turn from preliminary graphs of raw, uncontrolled plots of the data such as Figures 1 and 2 above to a regression implementation of the difference-in-difference approach that presents results in a simple table. This table will often focus on presenting the estimated coefficient of an interaction variable—i.e., the difference-in-difference coefficient—that captures how the outcome of interest changes on average before and after the reform in question and how that before–after change differs on average between the treatment and control groups. In other words, difference-in-difference papers will often collapse the entire exercise down to the reporting of a single
parameter of interest. And in the process, by collapsing everything into a binary before–after comparison, they lose the value that comes from a visual observation of how the outcome of interest truly trends over time.

With our analysis, our goal will be to maintain a visual, dynamic depiction of our results. In other words, even when we turn to our regression implementation, we will not abandon the spirit of the above graphs and collapse everything into a binary, before–after exercise. Rather, our goal will be to build on these motivating graphs and modify them to (1) estimate one trend line that directly depicts the differential in the two trend lines previously shown and (2) adjust that trend line for the potentially confounding influence of those factors that we want to control for in the regression.103

Arguably, the chief advantage of maintaining this dynamic approach is to be able to explore the pattern of results in the period prior to the reform in question—here, Arreola. Imagine that one estimates a classic, before–after difference-in-difference specification and derives a positive result that they summarize in a single parameter. Does this actually signify a positive effect of the reform? Possibly, but only under certain assumptions—i.e., that, but for the reform, the treatment group (here, non-white drivers) and the control group (here, white drivers) would have kept moving along parallel trends. To shed light on the reasonableness of that assumption, scholars often at least ensure that the treatment and control groups were moving in parallel in the period prior to the reform.104 A dynamic treatment of this regression

103 Technically, to estimate a dynamic regression of this nature is quite straightforward. A classic difference-in-difference regression regresses the outcome variable on an indicator variable for being in the POST period (i.e., after the reform), an indicator variable for being in the TREATMENT group (i.e., the group predicted to be impacted by the reform), an indicator variable for the interaction between being in the POST period and the TREATMENT group, and a series of control variables. The difference-in-difference estimate then comes from the estimated coefficient of that interaction term. To do this more dynamically, instead of using a single indicator variable to capture being in the POST period, we include a series of indicator variables for every month in our sample. We then interact all of these month indicators (except for one that will serve as our reference month) with the indicator signifying the TREATMENT group, which, in this case, is an indicator for being a non-white driver (or, alternatively, a Black driver). We then plot the estimated coefficients of these interaction terms. In essence, what these interaction terms do is allow us to observe how the stop rate for non-white drivers relative to white drivers evolves month by month, while teasing out the influence of the included control variables. Moreover, the whole exercise allows us to account for inherently fixed differences in stop rates between non-white and white drivers and inherently fixed differences in stop rates across individual months.

approach affords us that opportunity. Stated more broadly, perhaps the best way to truly assess whether Arreola appears to increase disparities is to observe the overall time path in racial disparities in stop rates and to visualize whether Arreola either led to a jump or an upward acceleration in the preexisting trend in racial disparities in stop rates. Collapsing everything down into a simple before-and-after comparison may mask important dynamics.

By way of preview, exploring such dynamics will turn out to be hugely important in the present context. Throughout the many alternative specifications that we estimate below, our analysis demonstrates a strong preexisting trend towards more stops of non-white drivers that materialized prior to Arreola. Further, we document no robust acceleration or jump in that preexisting trend coinciding with Arreola.

That said, while our focus will be on these dynamic graphs, there will be one occasion on which we estimate the more classic, single-parameter difference-in-difference approach. We acknowledge that the single-parameter approach is useful for summarizing the overall magnitude of the treatment effect size. In our case, since the point estimates of our findings will tend to suggest a near-zero effect of Arreola on traffic-stop rates, this exercise will be helpful in summarizing the magnitude of the confidence intervals of our estimates, as we discuss in greater detail below. In taking this approach, however, we will need to modify this classic approach to account for the preexisting differential trends in stop rates across races that we just discussed.

Before turning to our results, we make one final technical note regarding the functional form of our difference-in-difference approach. Given the nature of our outcome of interest—traffic stops—we will estimate a count model in our primary approach rather than an Ordinary Least Squares (OLS) specification. Primarily, we will estimate negative binomial regressions, given the presence of overdispersion in our data. The dependent variable will be the number of traffic stops associated with each unit of observation. The unit of observation will vary across the different approaches we take below (e.g., in our first approach, it will be a given county–month–race group). Nonetheless, this approach will still allow us to


105 For information regarding count models such as Poisson and negative binomial regression models, see generally A. COLIN CAMERON & PRAVIN K. TRIVEDI, REGRESSION ANALYSIS OF COUNT DATA 263–78 (2d ed. 2013).
normalize by population so that we can interpret our results in terms of the differential in the per-population rate of stops across race groups.\textsuperscript{106}

Scholars employing difference-in-difference methods often face design choices of this nature—e.g., OLS versus negative binomial—and of related natures—e.g., should one specify the outcome variable in levels, natural logs, or per-population rates, and so on. In the Online Appendix, we offer a more in-depth explanation behind the functional-form choices that we make in implementing our difference-in-difference design.\textsuperscript{107} To demonstrate the robustness of our findings, we also present results considering a range of permutations of these functional-form choices. For instance, we demonstrate that the conclusions gleaned from our negative binomial results generalize to OLS specifications that use log stop counts as the dependent variable, in addition to Poisson specifications.

Ultimately, in all but one of the specification permutations that we estimate, we find results similar to what we present below—i.e., we observe a trend towards more stops of non-white drivers that long predated the 2012 change in Washington’s constitutional standards, suggesting the change in law made no impact. The one permutation that we estimate that does produce results suggestive of a causal impact of \textit{Arreola} (with increases in the traffic-stop differential that postdated \textit{Arreola}) is the specification employed by Rushin and Edwards, which (based on our reading) estimates an OLS specification focused on stop levels (as distinct from logs) and that does not normalize stop counts by the relevant population.\textsuperscript{108} As we discuss in the Online Appendix, given the substantial baseline differences in stop counts across races, we are concerned that the functional-form approach considered by Rushin and Edwards does not track the ultimate policy question of interest, which is whether the \textit{Arreola} reform will be associated with a relatively greater increase in traffic stops of non-white drivers than it will of white drivers. This point aside, our analysis at least demonstrates the instability of that positive \textit{Arreola} effect.

3. \textit{Difference-in-Difference Results: County Controls}

In this first set of regression results that we present, we control for the county in which the stop takes place in order to address the above-stated concern over the racial composition of counties changing over time, combined with a concern over fixed differences in traffic-stop propensities

\textsuperscript{106} In particular, we include the adult population of the relevant unit of observation as an “exposure” variable in the regression so that we account for the inherent number of opportunities for the population of each unit of observation to be stopped.

\textsuperscript{107} See Online Appendix at 1550–60.

\textsuperscript{108} See Rushin & Edwards, \textit{supra} note 21, at 706.
Let us begin in Figure 3 by implementing this approach while using non-white drivers and white drivers as the racial groups of comparison. As above, we can think of these graphs as plotting a monthly trend in the difference between the stop rates of non-white drivers and white drivers. The relevant difference is normalized to zero in the month prior to Arreola, which will serve as the reference month. Positive values on the y-axis reflect a differential stop rate for the indicated month that is greater than the differential stop rate in the reference month, and vice versa for negative values on the y-axis.¹⁰⁹

To execute this approach, we organize our data such that the unit of observation is defined as a given county–month–race group. The dependent variable represents the number of stops of drivers in that county–month–race group. As above, we include the adult population count associated with that group as an “exposure” variable so that this stop count effectively entails a per-population stop rate. Moreover, as above, the regression will include separate indicator variables signifying each month and separate indicator variables signifying each racial group, along with a set of variables capturing interactions between each monthly indicator and the racial-group indicator. Again, to arrive at the time trend in the differential stop rates across races—our findings of interest—we plot the coefficients of these interaction terms. Importantly, in order to account for the concern identified over demographic swings that differentially affect counties, we will include a series of fixed effects for each county in this regression specification. By doing this, we can estimate the trend of interest while accounting for completely fixed differences in stop rates across each county. More specifically, what we are doing here is ensuring that we compare non-white and white stop rates within each county, not comparing the differential stop rates across counties. And we are then effectively averaging this within-county-stop-rate differential over all of the counties.

Moreover, the differential stop rate itself is interpreted in relative, not absolute terms—that is, in terms of how much larger or smaller the non-white stop rate is as a fraction of the white stop rate (given that we are estimating a negative binomial specification, whose coefficients can be interpreted as the log change in the outcome when the relevant explanatory variable increases by 1). The graph then depicts how this relative differential stop rate changes over time.

¹⁰⁹ To execute this approach, we organize our data such that the unit of observation is defined as a given county–month–race group. The dependent variable represents the number of stops of drivers in that county–month–race group. As above, we include the adult population count associated with that group as an “exposure” variable so that this stop count effectively entails a per-population stop rate. Moreover, as above, the regression will include separate indicator variables signifying each month and separate indicator variables signifying each racial group, along with a set of variables capturing interactions between each monthly indicator and the racial-group indicator. Again, to arrive at the time trend in the differential stop rates across races—our findings of interest—we plot the coefficients of these interaction terms. Importantly, in order to account for the concern identified over demographic swings that differentially affect counties, we will include a series of fixed effects for each county in this regression specification. By doing this, we can estimate the trend of interest while accounting for completely fixed differences in stop rates across each county. More specifically, what we are doing here is ensuring that we compare non-white and white stop rates within each county, not comparing the differential stop rates across counties. And we are then effectively averaging this within-county-stop-rate differential over all of the counties.

¹¹⁰ Moreover, the differential stop rate itself is interpreted in relative, not absolute terms—that is, in terms of how much larger or smaller the non-white stop rate is as a fraction of the white stop rate (given that we are estimating a negative binomial specification, whose coefficients can be interpreted as the log change in the outcome when the relevant explanatory variable increases by 1). The graph then depicts how this relative differential stop rate changes over time.
The key takeaway from this graph is quite simple. In short, this Figure does not present evidence to suggest that the *Arreola* decision caused a divergence in stop rates between non-white and white drivers. To be sure, at least when we focus on the point estimates, we do appear to observe that the degree to which non-white drivers are stopped (per non-white adult population) is growing relative to the degree to which white drivers are stopped (per white adult population). However, that trend appears to have started at the beginning of the sample period, several years prior to *Arreola*, and it does not appear to have accelerated or jumped as a result of *Arreola*. These findings simply do not suggest that *Arreola* contributed to a greater degree of racial disparities. Rather, disparities may be growing over this time period in Washington without a further contribution from the sanctioning of pretextual traffic stops.

While this upward trend is apparent from the point estimates, the standard errors around each monthly estimate are somewhat large in this approach, in part due to the taxing nature of this estimation strategy and the need to estimate a large number of monthly parameters given the available data. To get more statistical power, we estimate a trend line in Figure A12 of the Online Appendix that calculates differential stop rates annually as
opposed to monthly, thereby estimating far fewer parameters. As with the monthly graph from Figure 3, the point estimates from this annual-interval specification still suggest an upward trend in this differential stop rate over the whole time horizon—even before Arreola. With somewhat tighter confidence intervals, this annual-interval specification suggests a statistically significant jump in the non-white-to-white differential stop rate in the year prior to Arreola followed by increases in the differential rates after Arreola that are nonetheless not statistically distinguishable from zero. Again, these results are inconsistent with a story in which Arreola contributed to an elevation in racial disparities in traffic-stop rates.

In Figure 4, we replicate the analysis from Figure 3 but instead compare trends in the stop rates of Black and white drivers. As with Figure 3, we also show a counterpart in the Online Appendix (Figure A13) that produces this trend in annual intervals. Before Arreola, it looks as if the Black–white stop rate is trending downwards, and after Arreola, it looks like this differential rate is trending upwards. This pattern might indeed be suggestive of an effect of Arreola that increases the differential in traffic-stop rates between Black and white drivers. However, one key aspect of Figure 4 cuts against this interpretation. Over 2012—in the months leading up to the Arreola
decision—the Black–white stop differential is increasing (with the exception of the month before Arreola). As such, it could be that the trend towards more stops of Black drivers in the post-Arreola period is reflective of some phenomenon that predated Arreola and that was thus not caused by Arreola. In any event, in our attempts to account for other variables that are changing over time and that may also impact the Black-to-white stop-rate differential, the analysis that we set forth below will demonstrate that this initial observation of declining Black–white stop rates over the pre-Arreola period followed by increasing Black–white stop rates in the post-Arreola period does not hold up to further scrutiny. The analysis to follow suggests that the Black–white trend is similar to the non-white-to-white trend, with both generally increasing over the whole sample period, as opposed to in the post-Arreola period only.

4. Null Effects Versus Zero Effects

While the above graphs do not suggest that Arreola affected racial bias in traffic stops, this does not necessarily mean that the true effect of Arreola on racial disparities in stops is zero. Our estimates are just that—estimates. As with all estimates, they are subject to some degree of statistical noise. Accordingly, a more accurate description of our findings is that we identify a range of possible effects of Arreola on the differential in stops between non-white and white drivers, where this range is centered around zero. But we acknowledge there is some probability of a positive impact of Arreola on racial bias in stops. Consistent with best practices when presenting null effects, we will at least attempt to explain the magnitude of Arreola effect sizes that we can rule out with statistical confidence.

In doing so, this will be the one instance in which we will move away from presenting rich, dynamic graphs and try to characterize our findings in a single parameter. For these purposes, we collapse our findings down to a classic difference-in-difference estimate, whereby we simply estimate the average stop rate for non-white drivers relative to white drivers after Arreola relative to before Arreola. We can make this comparison by regressing the stop rate on an indicator for the non-white-driver group, an indicator for the post-Arreola period, and an interaction between these two indicators.

To be sure, when we estimate that classic difference-in-difference regression, we find a positive coefficient for the interaction term, specifically a roughly 5% increase in the relative non-white-to-white stop rate after Arreola. However, as Figure 3 shows, that positive effect is likely reflective of a seemingly linear upwards trend in the differential stop rate that predated Arreola. It is not likely caused by Arreola. The shape of Figure 3 suggests that the traditional difference-in-difference specification should be modified to account for race-specific linear trends in stop rates that transpire over the
full sample period. When we include such trends in the regression, the question that we ask effectively becomes: how does the stop rate for non-white drivers relative to white drivers increase in the post-Arreola period, where the non-white and white stop rates are each measured relative to the preexisting linear trends in such rates over time? This ensures that we would not falsely attribute a general trend towards relatively more non-white stops over time as being due to Arreola. Now, if Arreola happened to result in a one-time increase in stop rates for non-white drivers relative to white drivers, or if it happened to intensify the degree of the preexisting trend towards non-white stops, then this modified specification would capture such an effect.

Of course, a simple observation of Figure 3 already suggests neither an upwards bend in the differential trend line that occurs at the Arreola decision nor a one-time jump in differential stop rates. Accordingly, based on Figure 3, one would expect that estimating a simple before-versus-after difference-in-difference model that fits race-specific linear trends would produce an effect of Arreola that is close to zero. And in fact, this is exactly what we find. Specifically, as demonstrated by Column 1 of Table 2, this approach produces a point estimate for the Arreola effect of 0.006. This can effectively be interpreted as a 0.006 fractional change (or a 0.6% change) relative to the mean per-population stop rate in the data of 0.0128 — i.e., this point estimate suggests that Arreola increased the gap between non-white and white stops by a magnitude of about 8 stops per 100,000 people (0.6% of 0.0128).

Critically, this magnitude is both very small in and of itself and statistically indistinguishable from zero. In other words, we cannot say with statistical confidence that Arreola had any effect on stop rates. Even though this null effect is perhaps to be expected from the results presented in Figure 3, the advantage of the single-parameter approach is to help quantify the breadth of the possible effect size implied by our findings.

On this latter point, we find that the 95% confidence interval for the difference-in-difference coefficient estimated in Table 2 spans from (1) a

---

111 Specifically, we add to the basic difference-in-difference regression a linear trend term and an interaction between the linear trend and the non-white driver group. This specification will allow us to estimate the difference-in-difference effect while allowing for unobservable factors that drive a difference between the non-white and white stop rates and that evolve at a fixed, linear rate over the whole sample period. Scholars have characterized a modification of this nature as essentially identifying the reform effect under a "parallel trends in-trends" assumption. See, e.g., Naoki Egami & Soichiro Yamauchi, Using Multiple Pre-Treatment Periods to Improve Difference-in-Differences and Staggered Adoption Designs 12–13 (Sept. 15, 2021) (unpublished manuscript), https://soichiroy.github.io/files/papers/double_did.pdf [https://perma.cc/NZV2-SPEB] (explaining and illustrating the use of this linear term for modeling under the "parallel trends in-trends" assumption).

112 Again, the negative binomial coefficient signifies the log change of the relevant variable, which can be interpreted as a percent change. See Matthew A. Andersen, Calculating and Interpreting Percentage Changes for Economic Analysis, 1 APPLIED ECON. TEACHING RES. 25, 26 (2019).
To be clear, the point estimate discussed in the preceding paragraph—which implied an increase in the stop-rate difference of a magnitude of around 8 stops per 100,000 people—is notably closer to zero than these bounds and represents what we would expect to find from the Arreola decision. These bounds represent remote possibilities. What is helpful about this demonstration, however, is to illustrate that even the remote possibilities represented by the bounds of the confidence intervals are still relatively minor in degree.

Notably, the point estimate of the Arreola effect on racial disparities in stops from Rushin and Edwards—which suggests an increase in the non-white-to-white stop rate of over 10% of the mean stop count—falls beyond the confidence bounds of our analysis. In other words, not only does the point estimate from our study fall substantially closer to zero relative to that of Rushin and Edwards’s study, but our analysis can also reject with statistical confidence that the effect of Arreola is as large as that reported in their study.

In Column 2 of Table 2, we extend this bounding analysis to our review of Black versus white stops. The conclusion reached here is much the same, with small point estimates that are indistinguishable from zero and bounds of the 95% confidence interval reflecting remote possibilities that are themselves somewhat modest in size. Unfortunately, since Professors Rushin and Edwards did not estimate the effects of Arreola on the differential in stops of Black versus white drivers, we cannot situate their analysis within our confidence bounds.

---

113 As discussed in note 103, supra, we acknowledge that estimating confidence intervals is arguably problematic given that we are only drawing on one treatment state. In future work, we hope to draw from the experiences around 2012 in a number of control states to build a confidence interval through an approach similar to that proposed by Conley & Taber, supra note 98. This point does not necessarily undermine our conclusion that we fail to find evidence suggestive of an increase in racial disparities in stop rates following Arreola; however, future work in improving upon our standard-error estimates may improve our ability to infer that the effect is indeed small at best.

114 We find even tighter confidence intervals with the exercise undertaken in Table 2 when we include controls for driver race and sex. For instance, in the Black-versus-white stop comparison, we find that the upper bounds of the confidence interval suggest an increase in the Black–white stop-rate gap that is only 4% of the mean stop rate (or an increase representing only 5 out of 10,000 drivers).
TABLE 2: NEGATIVE BINOMIAL ESTIMATES OF EFFECT OF ARREOLA ON DIFFERENTIAL RATES OF TRAFFIC STOPS BETWEEN NON-WHITE (OR BLACK) AND WHITE DRIVERS (INCLUDING RACE-SPECIFIC LINEAR TRENDS)

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>After Arreola</td>
<td>−0.052 (0.038)</td>
<td>−0.065* (0.034)</td>
</tr>
<tr>
<td>Non-White Driver</td>
<td>−0.283*** (0.079)</td>
<td></td>
</tr>
<tr>
<td>Black Driver</td>
<td></td>
<td>0.073*** (0.104)</td>
</tr>
<tr>
<td>After Arreola x Non-White Driver</td>
<td><strong>0.006 (0.027)</strong></td>
<td></td>
</tr>
<tr>
<td>After Arreola x Black Driver</td>
<td></td>
<td>0.027 (0.031)</td>
</tr>
<tr>
<td>N</td>
<td>5,472</td>
<td>5,460</td>
</tr>
</tbody>
</table>

95% Confidence Interval for Estimated Difference-in-Difference Coefficient

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Column 1</td>
<td>[−0.046, 0.059]</td>
<td></td>
</tr>
<tr>
<td>Column 2</td>
<td>[−0.034, 0.087]</td>
<td></td>
</tr>
</tbody>
</table>

Note. * indicates statistical significance at 10%; ** indicates statistical significance at 5%; *** indicates statistical significance at 1%.

5. Missing Race Data

A critical challenge facing this exercise is posed by the fact that roughly 27% of the records in the Washington State Trooper database do not indicate a race of the driver (whereas approximately 57% of the records indicate a white driver, 3.5% a Black driver, 7% a Hispanic driver, 4.6% an Asian or Pacific Islander driver, and 1% a driver of other race). For those traffic stops in which race is not identified, none of the driver demographics are identified—that is, for roughly 27% of the stops, there is simply no record of the driver’s age, sex, or race.115

In one of their key analyses, Professors Rushin and Edwards assume that all of those with missing race fields are stops for non-white drivers—i.e., they assign as “non-white” all those drivers with race missing.116 They premise this assumption on the idea that officers may not be recording race to conceal their racial bias in stops.117 That assumption, however, seems inconsistent with the observation just made that stops with missing race codes provide no demographic information at all. Below, we will set forth an additional empirical marker that questions the reasonableness of this assumption. Nonetheless, alternatively, Rushin and Edwards demonstrate that the findings they present are robust to simply dropping observations with

115 More specifically, among those without race, 99.95% of those stops do not report driver age and 99.9% do not report the sex of the driver. See Data, supra note 93.
116 Rushin & Edwards, supra note 21, at 669.
117 Id.
missing race, an approach that we find preferrable to their other approach and one of those that we take in our own analysis. Overall, our investigation into this issue builds on Rushin and Edwards’s work and will take several approaches in attempting to ascertain whether the presence of stops with missing race biases us against finding evidence of a true effect of Arreola.

At the beginning of this important discussion over missing race codes, let us clarify what we have already done in Figures 3 and 4 to begin to address this problem. When estimating the regressions underlying these Figures, we have taken one of the common empirical approaches to address missing data and have provided for a third race group in our data structure—non-white, white, and missing-race—while also including an indicator variable in the regression for this missing-race group, along with a set of interactions between this missing-race indicator and each sample month. With this approach, we are controlling for how the trends in the differences in non-white to white stops that we observe within counties—our key parameters of interest—may correlate with the time trends in stop rates among drivers with missing race that we also observe. That is, hypothetically, if the data tend to show that non-white stops fall relative to white stops when we also happen to see growth in the number of stops with missing race codes, then this regression approach can control for such an effect. We note, however, that the results turn out to be almost identical if we do not include these missing-race controls and simply drop those stops with missing race codes entirely from the analysis.

In the remainder of this Section, we nonetheless acknowledge the possibility that this missing-race-control approach may not sufficiently account for a bias that may result from missing race codes. Accordingly, we now turn to setting forth various empirical markers designed to diagnose whether we should indeed be concerned with such a bias.

To begin this discussion, recall that, thus far, our results have failed to produce evidence to suggest that Arreola contributed to an increase in racial bias in traffic stops. Might a positive effect of Arreola nonetheless exist? And might so many observations with missing race data prevent us from finding this positive effect? At the outset of thinking about this question, we of course acknowledge that if we could assign a race field to those with

---

118 See, e.g., Jason Abrevaya & Stephen G. Donald, A GMM Approach for Dealing with Missing Data on Regressors, 99 REV. ECON. & STAT. 657, 657 (2017) (describing the inclusion of an additional dummy variable to indicate when data are missing).

119 Recall that the data are organized at the county-by-month-by-race-group level. Here, we simply allow for a missing-race group within this structure. For the purposes of structuring our exposure variable, the population that we assign to the missing-race group is the total population for the relevant county–month.
missing race data, we could achieve even better precision in our analysis and even further tighten the confidence intervals discussed previously. But of greater relevance to our present inquiry is not precision but whether we should have reason to be concerned that our previous analysis has been biased by not having this information—and biased in a way that may be masking a true Arreola effect.

We can envision two scenarios in which missing race may be imposing such a bias. First, it could be that non-white drivers are inherently more likely to have a missing race code than white drivers. In other words, out of the group of approximately 27% of stops with missing race codes, more of them may in fact be non-white rather than white relative to the 73% of stops in which we can observe race codes. This fact alone actually would not be problematic; however, if one combined this fact with an increase in the rate of stops with missing race codes in the post-Arreola period, it could be true that we are not capturing all of the increase in non-white stops relative to white stops that is occurring after Arreola.

Alternatively, it could be that the overall mix of non-white and white drivers among the 27% of stops with missing race codes equals the mix of non-white and white drivers in the 73% of stops in which we can observe race (over all sample years). Even if that is the case, a problem could still be posed by the possibility that the proportion of non-white drivers among the group of drivers with missing race codes is higher in the post-Arreola period relative to the pre-Arreola period. A hypothetical scenario of that nature could also account for an increase in the number of stops of non-white drivers relative to white drivers after Arreola relative to before that we are not accounting for.

In the analysis to follow, we will take several steps to confront this question and to assess whether these two scenarios may in fact be confounding the conclusions suggested by the above Sections.

To begin, consider the first scenario just raised. There are two components to this concern: (1) that the missing-race group has a higher percentage of non-white drivers than the group for which race can be observed and (2) that we see a trend towards more stops with missing race that begins after Arreola. Let us take these in turn. First, it is ostensibly unknowable whether the missing-race group is relatively more non-white than the group with observable race. However, we can at least look at other characteristics of traffic stops that we can observe—e.g., the location of the stop (which is relatively fine-grained with roughly 7,500 specific locations recorded throughout the state in the data), the time of day of the stop, and the sex and race of the officer—and use them to predict a likelihood of the driver being non-white for all stops that have values for these other characteristics,
whether or not they have values for the driver’s race. By drawing on other stop characteristics to form driver-race predictions, we find only a very slightly higher likelihood of being non-white among the missing-race group (with a predicted likelihood of being non-white equal to 30.26%) compared with the observable race group (29.64%). This difference alone may be minor enough not to worry about this scenario in the first place, but again even if this difference were larger, it would only be a concern if we also saw a growth in missing-race stops that emerged in the post-Arreola period. Do we find this?

In Figure A2 of the Online Appendix, we set forth a monthly trend in the across-county average number of stops with missing race. From two years before and after Arreola, there is no apparent trend in the number of stops with missing race, just monthly noise. There is a notable jump up in the number of stops with missing race codes in the period of time after two years post-Arreola. One might be concerned that Arreola has a lagged impact on the number of stops of non-white drivers and this increase in the count of stops with missing race at the end of this sample is preventing us from observing this effect. However, this observation is not limited to only the post-Arreola period, cutting against the possibility that this increase in missing stops coincides with Arreola itself. That is, we also observe that in the period of time before two years pre-Arreola, we see sharp increases in the counts of stops with missing race. As such, if non-white drivers indeed have relatively more missing-race events recorded than white drivers, any correction that we would make for this given the trend in missing-race counts that we observe would produce an increase in non-white stops relative to white stops over the whole sample period—before and after Arreola. In other words, even if non-white drivers are relatively more likely to be overrepresented in the missing-race data and we tried to correct for that, we

---

120 Specifically, to form these predictions, we consider the sample of stops with information on driver race and the other variables of interest—e.g., stop location—and estimate a probit model of the incidence of the stop being associated with a non-white driver (or Black driver, alternatively) on variables indicating these other characteristics of the stop. We then form predicted values of driver race based on the coefficients from this probit specification, but we form these predictions over the whole sample—at least the whole sample that has information on the predictive variables. The predicted values from this specification allow us to form a prediction of the likelihood that the stop is of a non-white driver based on the observable values for these various characteristics. Addressing the problem of missing data through imputation exercises of this nature—i.e., estimating missing values based on other available data—is a common approach taken in empirical analyses to confront missing data. See, e.g., Abrevaya & Donald, supra note 118, at 657 (describing “imputation,” a method that involves estimating missing data using other available data).

121 This finding further cautions against the primary approach taken by Rushin and Edwards of assuming that all stops with missing race fields are for non-white drivers and lends support to the secondary approach that Rushin and Edwards take, which drops observations with missing race. Rushin & Edwards, supra note 21, at 669.
would likely just reinforce the conclusions reached above—that stops of non-white drivers relative to white drivers are increasing over the whole sample period and do not appear to be caused by *Arreola*.

Let us now move on to assess the second scenario posed above—i.e., the possibility that, whatever the overall share of non-white to white drivers is among the missing-race stops, the composition of this missing-race group is nonetheless trending towards relatively more non-white (or Black) drivers relative to white drivers in the post-*Arreola* period. We attempt to shed light on this possibility through a straightforward imputation exercise. In short, we will use other available information from the stop—as above—to form the predicted likelihood that the stop involved a non-white driver.\textsuperscript{122} For the purposes of tractability, we then move beyond just predicting the likelihood of being non-white but also predict in a binary sense which of these missing-race stops are for non-white drivers and which are for white drivers. By forming predicted race assignments in this binary sense, we will afford ourselves a straightforward means of comparing—via graphs similar to those already presented—the counts of stops between the group of drivers with missing race codes that are predicted to be non-white and the group of drivers with missing race codes that are predicted to be white.\textsuperscript{123} If this trend happens to demonstrate an increase in this particular differential that coincides with *Arreola*, it would implicate a concern that the missing-race data are biasing our analysis against observing a true positive impact of *Arreola*.

In Figure 5, we present results from this exercise, demonstrating the monthly trend in the differential stop rate for drivers predicted to be non-white relative to drivers predicted to be white among those stops with missing race codes. This trend is flat over the whole sample period, easing concerns that missing race data may be precluding us from documenting the positive effects of *Arreola* predicted by the dominant theory. In Figure 6, we repeat this exercise for drivers with predicted-Black race and predicted-white race among the set of stops with missing race codes. Again, we find no

\textsuperscript{122} See supra note 120 for information on this prediction exercise.

\textsuperscript{123} To be able to form these binary assignments, however, we need to make a normalization assumption—though it is an assumption that we have already made when describing this scenario. That is, we assume that the overall percentage of non-white drivers among the group of drivers with missing race codes is the same as the percentage of non-white drivers among the group of drivers with observable races. With our predicted likelihoods of being non-white, we can then order all of the missing-race drivers in order of these likelihoods and assign them into the non-white group in the order of this likelihood until this overall assumed percentage is met. To reiterate, this is simply a normalization exercise. Whatever the overall percentage of non-white drivers among the missing group happens to be, the purpose of this exercise is just to see how the composition of this group is trending over time—i.e., to see if the group of missing-race stops is trending (in a way that coincides with *Arreola*) towards more stops with characteristics that would lead one to predict they are likely associated with non-white drivers. On a final note, we performed all of these prediction exercises on a county-specific basis.
markers that suggest an increase in stops of predicted-Black drivers relative to predicted-white drivers among this missing-race group coinciding with *Arreola*, which would have raised concerns that we are biased away from finding an *Arreola* effect.

**Figure 5:** Trend in Differential Stop Rate of Predicted-Non-White Drivers and Predicted-White Drivers Among Stops with Missing Race Codes
6. Other Control Variables

Thus far, we have shown that there does not appear to be an increase in the rate of non-white-to-white (or Black-to-white) stops that arises in the post-Arreola period. So far, we have controlled for the county of the stop and for fixed differences in stop rates across months and across races. However, it is possible that some other characteristics of the underlying non-white (or Black) population are also changing over time in a way that could confound this analysis and mask a true effect of Arreola. For instance, hypothetically, one may be concerned that the population of non-white drivers is becoming relatively older post-Arreola. This may be concerning if older drivers are stopped at lower rates. That is, it could be that Arreola in fact increased the propensity to pull over a given non-white driver, but it is hard to observe that outcome based on our results thus far since we are hypothetically also observing a shift towards older non-white drivers.

Fortunately, we can address this concern—and a related concern based on the recorded sex of the driver—given that we also observe these two characteristics within the Washington State Trooper database. In Figures 7 and 8, we replicate Figures 3 and 4 from above but now estimate the differential trend in the rate of stops of non-white relative to white drivers (and, alternatively, Black relative to white drivers), but do so while
controlling in the underlying regression for the recorded age and sex of the driver.¹²⁴

FIGURE 7: DYNAMIC DIFFERENCE-IN-DIFFERENCE ANALYSIS: ESTIMATED TREND IN
DIFFERENCE BETWEEN NON-WHITE-TRAFFIC-STOP RATE AND WHITE-TRAFFIC-STOP RATE,
INCLUDING CONTROLS FOR DRIVER AGE AND SEX

¹²⁴ Specifically, to achieve this, we arrange the data at the county–month–race-group–age-group–sex-group level (e.g., there would be a unit of observation for non-white males aged twenty to twenty-four in King County in January 2013). We then estimate the same specification mentioned above, but we also include control variables for the different age and sex groups. Necessary for this approach is also population data at this level of aggregation, which we obtain from the American Community Survey (available at [https://data.census.gov/mdat/#](https://data.census.gov/mdat/#)). In this sense, we are determining these differential trends between non-white and white drivers within different demographic groups and then effectively averaging over all such demographic groups. In this way, we ensure that differences in stop propensities across age groups or across sexes are not contributing to our findings.
The results continue to dispel any notion that Arreola causally contributed to a rise in non-white stops. If anything, as before, we simply see a general trend over time towards more non-white stops, with the Arreola decision having no apparent impact on that trajectory. This appears to be the case both when comparing non-white and white driver trends and Black and white driver trends.

In Figures A9 and A10 of the Online Appendix, we take this controls analysis one step further and organize the data at the county–month–driver-race–driver-age–driver-sex–officer-age–officer-sex level, allowing us then to ensure that we compare the rate of stops of non-white and white drivers within officer race and sex groups. This addresses potential concerns that the trends presented thus far may be confounded, for instance, by changes in the racial composition of the officer workforce over time, which is concerning to the extent that officers may differ across race in the degree of their racial bias in initiating stops. We continue to estimate similar patterns in the data by incrementally adding officer controls of this nature.

**D. Triple-Differences Analysis: Night-Versus-Day Variation**

Even after controlling for the possibility of demographic shifts in the county of residence and in the age and sex of drivers, as we have done above,
one may still be concerned with yet other characteristics of drivers changing differentially for white and non-white drivers in the post-Arreola period. That is, it could still be the case that Arreola led to an increase in the non-white-to-white stop differential, but we are unable to discern it due to some shock that we cannot observe and thus cannot control for.

In this Section, we will consider another quasi-experimental strategy to address this concern. Drawing on other studies in racial bias in traffic stops, we start with the prediction that officers exhibiting racial bias in traffic stops will be able to act on such biases during the daytime, but not during nighttime hours (or at least not as effectively during nighttime hours). This prediction motivates adding another layer of control—i.e., another layer of differentiation—to the strategy that we have employed above. Thus far, our analysis has stemmed from the idea that if Arreola affected stops, we would see such an effect on non-white drivers in the post-Arreola period. This, in turn, motivated us to use white drivers as one layer of control and to use the pre-Arreola period as another layer of control. A “veil-of-darkness” prediction allows us to add yet another dimension to the analysis and use nighttime stops as a third layer of control.

In essence, we will build on the difference-in-difference strategy from above to estimate a “triple-differences” specification. In this approach, we will estimate the difference between day and night in the difference-in-difference calculation from above. If Arreola is expected to increase stops for non-white relative to white drivers, and if one would expect this effect to be larger during daylight hours, then one would expect a positive triple-differences finding (when viewing nighttime as the control).

This veil-of-darkness methodology effectively leaves us with a within-race treatment and control group. Why is that especially valuable? Consider a hypothetical threat to the analysis thus far: We may have not observed an expansion in the non-white-to-white stop rate after Arreola because the non-white population experienced a negative income shock in the post-Arreola period that left them relatively less able to buy a car. Our close-to-zero point estimates from Table 2 could perhaps be explained by a combination of a true increase in racial bias resulting from Arreola that is nonetheless offset in the aggregate by this hypothetical negative income shock to the non-white

population. But, this income shock would be expected to lower the non-white stop rate both at night and during the day. By taking our estimated time trend in the non-white-to-white stop rate and further differencing that trend between night and day—i.e., via a triple-differences estimate—we would effectively take our difference-in-difference finding from above and cleanse it of unobservable shocks of this hypothetical nature. That is, we can identify the effect of *Arreola* even while accounting for the possibility that something we cannot observe is driving up or driving down the stop rate of non-white drivers in the period following *Arreola*. This is the power of the triple-differences strategy.

In executing this strategy, we continue to avoid collapsing the inquiry into a simple before-after analysis—i.e., we avoid estimating just a single triple-differences parameter. We again want to take a more dynamic approach and track these differences over incremental time periods leading up to and following *Arreola*. However, given that our specification is a bit more intricate now and involves a notably greater number of parameters to estimate, we present these triple-differences results using annual increments rather than monthly increments.\(^{126}\) So, year by year between 2010 and 2015, we show how the differential stop rate between non-white and white drivers—its own differentiated between day and night—trends. If *Arreola* caused an increase in stop rates for non-white drivers, we would expect to observe a positive value of this double-differential emerge in the 2013–2015 period, which is the post-*Arreola* period. We present results for this exercise in the next two graphs, first comparing non-white and white drivers and then comparing Black and white drivers.

\(^{126}\) The conclusions we reach, however, are robust to using a monthly approach. To formalize this dynamic triple-differences estimation, we regress the number of stops on a series of triple-interaction terms—i.e., an indicator for non-white drivers interacted with an indicator for daytime hours interacted with an indicator for 2010, and so on for each sample year, leaving out the year leading up to *Arreola* to serve as a reference year. The regression specification also includes all constitutive terms of these interactions—i.e., all relevant double-interaction terms along with the indicator variables for each month, for daytime hours and for non-white drivers. Naturally, to pull off this specification, we organized these data at the county-by-year-by-race-group-by-day–night level. For the exposure variable, we continue to use the relevant year–month–race population count and do not use a separate exposure variable for day versus night (as we have no natural candidate for such purposes). We also confirm that our results are essentially identical when we do not include an exposure variable. The general triple-differences methodology is motivated by Jonathan Gruber. See Jonathan Gruber, *The Incidence of Mandated Maternity Benefits*, 84 Am. Econ. Rev. 622, 627 (1994).
Figure 9: Dynamic Triple-Differences Analysis: Estimated Annual Trend in Difference between Non-White-Traffic-Stop Rate and White-Traffic-Stop Rate during the Day Versus at Night

Figure 10: Dynamic Triple-Differences Analysis: Estimated Annual Trend in Difference in Traffic-Stop Rate Between Black Drivers and White Drivers during the Day Versus at Night
With respect to Figure 9, showing non-white versus white stop-rate trends (day versus night), we do see higher average rates after 2012 than before. However, the double-interaction effects—non-white versus white, further differentiated by day versus night—had already been trending upwards prior to Arreola. As can be observed, the estimated coefficient for 2010 is statistically distinguishable from its 2012 level, demonstrating a statistically significant and large pre-Arreola trend. As such, this graph is inconsistent with a story in which Arreola—i.e., the relaxation in the prohibition of pretextual stops—caused officers to increase racial bias in traffic stops, as the dominant theory predicts. If anything, the fact that this increasing pre-trend appeared to flatten out after Arreola suggests an effect in the other direction.

As mentioned above, we are inclined to focus on a three-year observation window in our preliminary difference-in-difference approaches given that we possess only one treatment state in this natural experiment and given concerns over unobservable Washington-specific shocks that may inevitably arise at some point in the future. But we do consider longer postreform observation windows in instances in which we use more advanced approaches designed to address such unobservable shocks. This day–night triple-differences approach is one of those instances. Accordingly, in Figures A16 and A17 of the Online Appendix, we present counterparts to Figures 9 and 10 that extend this window by three additional years—i.e., through 2018. This expanded window does not alter the conclusions we reached above. We do observe a sharp increase in the non-white-to-white differential stop rate in 2016, but this spike is not long-lived and systematic, with the differential returning to close to zero in 2017 and 2018. And we actually observe a decline in the Black–white differential stop rate in 2017 and 2018. Altogether, there is no evidence that Arreola led to delayed worsening of racial bias in traffic stops.

To conclude, we reiterate that the veil-of-darkness method is designed to alleviate concerns over unobservable race-specific shocks that would affect traffic stops during the day and night alike—e.g., it can account for the income shock hypothesized above. But what if there are reforms that were passed around the time of the Arreola decision that may also affect an officer’s racial bias in initiating traffic stops? This veil-of-darkness methodology would simply not allow us to separate an Arreola effect from the effect of these other reforms. Why? Because if these other reforms hypothetically affect an officer’s racial bias, they too will ostensibly have a stronger effect during the day than during the night.

We emphasize this point as there is one reform of special concern here—i.e., Washington’s legalization of recreational marijuana that became
effective in the same exact month that the Arreola decision came down.\textsuperscript{127} Professors Rushin and Edwards motivated their day-versus-night triple-differences approach apparently in part as a means of separating an Arreola effect from a marijuana legalization effect.\textsuperscript{128} However, per the point just made, if marijuana legalization resulted in a change in racial bias in the decision to pull over a driver, simply comparing stops in the day versus the night would not isolate the effects of marijuana legalization from those of Arreola. After all, both reforms could impact an officer’s bias in the stop decision, leading to a prediction of a stronger effect in the day over the night for both reforms. Accordingly, we now turn to a novel approach not yet considered in the literature that may also allow us to confront this particular challenge and to help disentangle the effects of marijuana legalization from the effects of a change in substantive constitutional standards bearing on pretextual stops.

\textbf{E. Triple-Differences Analysis: Washington Versus Colorado}

As just stated, if the legalization of recreational marijuana affected officers’ racial motivations in initiating a traffic stop, the contemporaneous developments of Arreola and marijuana legalization will naturally leave it difficult to disentangle the effects of these separate developments, even when employing the day-versus-night differentiation. But why might one expect marijuana legalization to also affect an officer’s bias? We can envision a few possible stories.

For instance, marijuana legalization could leave officers more inclined to stop non-white relative to white drivers to the extent that legalizing marijuana created an increased scope for stops based on driving while under the influence of marijuana\textsuperscript{129} and to the extent that officers perceived non-white drivers as more likely to commit this offense. On the other hand, legalizing marijuana reduces one basis for initiating a stop—i.e., to search for marijuana in the car. To the extent stops of that nature were racially biased, legalizing marijuana could theoretically reduce the scope on the margin for racial bias in initiating traffic stops.\textsuperscript{130} This latter possibility is particularly concerning because it could mean that Arreola may have


\textsuperscript{128} Rushin & Edwards, supra note 21, at 683, 690–91.

\textsuperscript{129} See WASH. REV. CODE § 46.61.502(1)(b), (5)–(6) (2022).

\textsuperscript{130} Pierson et al., supra note 17, at 736, did show that the racial gap in searches was reduced after marijuana legalization. However, that finding was focused on searches and not initial stop decisions, which are the focus of this Article.
exacerbated racial bias in traffic stops, but we simply cannot observe that marginal effect because it was washed away by a countervailing negative effect of legalizing marijuana.

Fortunately, we are able to take steps to confront this concern by comparing the experiences in Washington with the experiences in Colorado, a state that legalized recreational marijuana use with an effective date for that legal change only days away from the corresponding effective date in the state of Washington (a date that again was in the same month as the *Arreola* decision). In essence, in order to address the concern just raised, we can look to the experiences of Colorado around this time and see how the non-white to white stop differential changed in Colorado before and after 2013. In Figures 11 and 12, we show results of this nature, effectively replicating the Washington analysis from Figures 3 and 4 but using Colorado data.

---

**Figure 11: Dynamic Difference-in-Difference Analysis: Estimated Trend in Difference Between Non-White-Traffic-Stop Rate and White-Traffic-Stop Rate in Colorado**

---

131 Amendment 64 to the Colorado constitution, ratified on November 6, was proclaimed into the Colorado constitution on December 1, 2012. Marijuana Enforcement Division, *Permanent Rules Related to the Colorado Retail Marijuana Code*, COLO. DEP’T OF REVENUE (Sept. 9, 2013), https://www.colorado.gov/pacific/sites/default/files/Retail%20Marijuana%20Rules,%20Adopted%200913,%20Effective%20101513%5B1%5D_0.pdf [https://perma.cc/ZEB7-TFXT].
Interestingly, this Colorado analysis suggests, if anything, that legalizing marijuana may have halted a preexisting downward trend in non-white stops relative to white stops (and in Black relative to white stops). In other words, marijuana legalization may have increased racial bias in traffic stops. Again, our concern with the Washington analysis was that perhaps marijuana legalization decreased racial bias, which is why we could not detect an increase in bias after Arreola (as predicted by the dominant theory). But that concern does not seem to bear out according to Figures 11 and 12.

To elaborate, let us assume that the Colorado marijuana legalization experience can inform on the marginal effects of legalizing marijuana in Washington, an assumption aided by the fact that the marijuana legalization occurred nearly contemporaneously in both states. Proceeding with this assumption, if we were to net out the marijuana effect implied by the experiences in Colorado (see Figures 11 and 12) from the pre-post-Arreola effects in Washington, one might conclude that the non-white-to-white differential in Washington in the post-Arreola period should be even smaller than that depicted in Figures 3 and 4 above. This only further cuts against any inference that Arreola led to an increase in the rate of non-white relative to white stops.

We next formalize this comparison of the experiences of Washington and Colorado before and after the end of 2012 by estimating a triple-
differences specification. The goal with this strategy is again to estimate the effect of Arreola on racial bias in traffic stops while netting out the influence of legalizing marijuana on racial disparities in stops. Methodologically, this approach is similar to the veil-of-darkness approach, except that instead of using day versus night as the third dimension of differentiation, we compare stop rates in Washington and Colorado. The key assumption of this approach, of course, is that marijuana legalization will impact Colorado and Washington in the same manner, at least as it relates to racial bias in traffic stops. As with the day-versus-night approach, we will again use annual as opposed to monthly increments for this analysis, given the much greater number of parameters involved in the estimation.

**Figure 13: Dynamic Triple-Differences Analysis: Estimated Trend in Difference Between Non-White-Traffic-Stop Rate and White-Traffic-Stop Rate in Washington Relative to Colorado**

![Graph showing differences in traffic stop rates between non-white and white drivers in Washington and Colorado over years 2010 to 2015, with confidence intervals and point estimates marked.]

In Figures 13 and 14, we present the triple-differences results in a dynamic graph (with Figure 13 focusing on the non-white-to-white comparison and Figure 14 the Black–white comparison). Consider first Figure 13. With this graph, we can see how the rate of stops of non-white drivers relative to white drivers in Washington relative to Colorado evolves

---

132 See supra Section IV.D.

133 Stated differently, this approach assumes that other than Arreola, any race-specific shocks to stop rates that may hit in the post-Arreola period will affect Colorado and Washington alike (in percent terms).
year by year. If Arreola increased racial bias, we would expect to see a jump in this double-differenced measure after Arreola—i.e., after the end of 2012. As is evident from Figure 13, however, we do not see such an increase. Rather, prior to Arreola, there is already a statistically significant preexisting trend towards greater non-white-to-white stop rates in Washington relative to Colorado. If Arreola contributed further to racial bias in stops, we would expect that trend to tilt even more upwards or to jump to a new level. Instead, the trend turns downwards slightly (which is arguably consistent with the implications of the trend break suggested by Figure 11).

The results from the Black-to-white stop-rate analysis from Figure 14 show less evidence of a preexisting trend over the whole pre-Arreola period, though we do observe a large jump from 2011 to 2012. As such, even if one views the pattern of point estimates as an increase in the post-2012 period—which itself is arguably unclear from this Figure—that increase would have started to materialize in 2011.

Altogether, these results are simply not supportive of the dominant theory that weaker constitutional protections against pretextual stops intensify racial bias in traffic stops. This conclusion is not altered when we consider the possibility of a lagged response and incorporate a longer post-Arreola reform window, as we demonstrate in Figures A14 and A15 of the Online Appendix. In the case of the non-white-to-white differential, we do
observe further increases in this differential in 2016 and 2017, but again, this longer dynamic picture continues to suggest that we may simply be observing a trend that predated *Arreola*.

**F. Seattle Police Department Data Analysis**

We now address one final empirical concern. We have conducted this analysis thus far using data from the Washington State Patrol. However, one may be concerned that we investigated the impact of a change in the law bearing on pretextual stops in a setting in which one might have predicted a weaker impact in the first place. After all, even in the pre-*Arreola* period during which pretextual stops were prohibited in Washington, it would perhaps be hard to challenge the decision of a state trooper since much of their job entails regulating traffic. Perhaps the change in the law would be expected to more meaningfully change the outcome in cases involving other types of officers—i.e., officers perhaps primarily focused on nontraffic offenses. In the Online Appendix, we address this question by turning to traffic-stop data from the Seattle Police Department (SPD). Though these data are limited in certain respects relative to the state trooper data, our analyses of the SPD data are consistent with that set forth above, further reinforcing our conclusion that *Arreola* did not appear to increase racial disparities in traffic-stop rates.

**V. IMPLICATIONS**

**A. *The Exclusionary Rule***

Our empirical analysis can be seen in one of two lights. One can arguably view our analysis as demonstrating that the 2012 substantive change in Washington constitutional law had *little to no impact* on racial disparities in traffic-stop rates. At the very least—being mindful of the inherent empirical challenges facing this exercise that may leave us slightly less inclined to speak so definitively—one can view our analysis as finding *no evidence to support the claim* that the *Arreola* decision increased racial

---

134 Taking those five post-*Arreola* years together in the non-white-to-white differential graph, it appears that the slope of the post- *Arreola* trend in this stop-rate differential is no greater than the corresponding slope in the pre-*Arreola* period. If anything, the post-*Arreola* slope is less than in the preperiod. In the Black–white stop-rate Figure, we continue to see an increase in the point estimates in 2016 and 2017, though the differential in 2017 relative to the reference period still remains statistically indistinguishable from zero, and the largest jump that we observe in this differential occurs in the year prior to *Arreola*. This raises doubts that this trend in the point estimates may be reflective of an effect of *Arreola* itself.

135 See Online Appendix at 1571–73.
disparities in stops and thus no evidence to support the claim that Whren may contribute to racial disparities in stops.

In Part II, we theorized that Arreola’s lack of effect on officer behavior might be due to certain inherent limitations in the enforcement of substantive constitutional standards, including (1) the difficulty in applying the Ladson prohibition against pretextual stops given the difficulty in establishing pretext and (2) the small likelihood that an officer’s decision to initiate a stop will be impacted by the remote possibility that any evidence found during a pretextual stop will be suppressed, especially when the officer may have motivations beyond investigating crimes—e.g., hitting stop quotas. From this perspective, one can view our empirical analysis and the null effects that we present as providing support for the theory that the exclusionary rule is ineffective in shaping officer behavior in the context of traffic stops.

Of course, it would be inappropriate to draw broader conclusions about the exclusionary rule in other contexts. For instance, the exclusionary rule is probably important in shaping a homicide detective’s behavior when she is deciding whether to enter a suspect’s house. But where some combination of the above two factors—low likelihood of punishment and noninvestigative purpose—is present, we should be at least skeptical that the exclusionary rule will effectively encourage Fourth Amendment compliance.

B. Reducing Racial Bias in Traffic Stops

Our results are helpful for policymakers and others interested in reducing racial bias in traffic stops. Most obviously, our results caution that strategies focused on implementing a Fourth Amendment right—or state constitutional equivalent—against pretextual stops might not be as impactful as many assume. The same goes for a statutory equivalent that provides an exclusionary remedy as the sole deterrent measure.

It might be worth exploring the combination of a tailored test for impermissible motive with a more effective deterrent than the exclusionary rule. Our theory for why a Fourth Amendment-type solution would fail to provide a reliable and implementable test for pretext was predicated on two types of concerns—both of which might be avoided through extraconstitutional policy changes. First, judges might be unreliable in applying any test against the police in the context of a suppression hearing. Second, a test for pretext will be difficult to apply with confidence to an individual stop. These problems might be avoided by punishing officers based on their aggregate performance, not individual stops, and having the system implemented by police departments or other agencies, not courts.

136 See supra Section I.C.
Perhaps precincts could track and analyze their officers’ stops. Even though it is difficult to tell whether a given stop was made for good or bad reasons, longer term trends of racially disparate stops can be a red flag that triggers certain disciplinary responses.

There are, however, significant challenges even with a system like this. Skeptics will point to the “benchmarking” problem. If the underlying rates of traffic violations are different across races, we should expect stops to be disproportionate. As a result, it is difficult to know that an officer is doing something wrong simply because her stops are not proportionate across race. We suggest several possible approaches to this challenge.

First, while this benchmarking problem may present a real challenge when looking at aggregate counts of stops, it may be more feasible to detect an officer’s inappropriate, disproportionate treatment of drivers by race by focusing more specifically on certain characteristics of stops. For instance, one can look at the distribution of officers’ stops of Black drivers across the different speed-violation ranges—e.g., 0–5 miles per hour (mph) over the speed limit, 5–10 mph over the limit, etc. If one finds that this distribution is heavily concentrated in the low-violation ranges, at least relative to officers’ corresponding distribution for white drivers, this would cast doubt on claims that the greater overall volume of stops for Black drivers is due to their elevated driving speeds. If that were true, one would at least expect that these violation-speed distributions would be roughly similar across driver races. So, if the higher relative volume of Black stops is concentrated on minor offenses, this pattern is suggestive of both pretext and bias by officers.

Relatedly, if for a given reason and type of stop—e.g., driving 5–10 mph over the limit—an officer searches Black drivers at a higher rate than white drivers, this will naturally raise red flags regarding racial bias by the officer. Further, an officer’s “hit rate” across races—the proportion of searches that actually turn up contraband can be helpful in figuring out whether an officer’s behavior is acceptable. These latter metrics may be less indicative of pretext but are certainly suggestive of bias.

What types of remedies should be considered when these metrics raise red flags of bias or pretext? More effective remedies—i.e., more effective than the exclusionary rule—might include taking away scheduled pay raises

---

137 Grogger & Ridgeway, supra note 125, at 878 (discussing that the “key problem” when testing for racial disparities in traffic stops is the “benchmarking problem,” referring to the challenge in identifying a “risk set” of behaviors against which to compare “the racial distribution of traffic stops” that occur).

138 Id.

139 Rushin & Edwards, supra note 21, at 725–26.
or promotions for officers who have trended towards racially biased stops.\textsuperscript{140} Pay and promotion are likely to be more important to a given officer than the admissibility of evidence at some future trial. But it is also important that any solution ensures that superior officers are encouraged to actually follow through—higher-ups should be rewarded rather than punished for sanctioning subordinates.\textsuperscript{141} Independent oversight may also be helpful in this regard.\textsuperscript{142} And it is important that the punishment is regularly and evenly applied with evident procedural fairness. Otherwise, the sanction might backfire and actually increase noncompliance.\textsuperscript{143}

Ultimately, even though these solutions might have a better chance at success in reshaping officer behavior than a federal or state constitutional rule change, they face significant obstacles because they still operate in a world in which officers have tremendous discretion about whether or not to make a given stop. Once we accept that officers can stop some people who commit a given traffic offense and not stop others who commit that same offense, it will be incredibly difficult to police and punish bad stops after the fact. Thus, it is possible that the most fruitful avenue for reducing racial bias in traffic stops is to simply ex ante reduce officer discretion in making traffic stops.

For example, as technology advances, traffic laws could be enforced remotely and automatically—obviating any human discretion about whom to stop.\textsuperscript{144} And even without affirmative government adoption of technology, as self-driving cars are adopted, we might expect traffic violations to largely disappear.\textsuperscript{145} In that case, the scope for police discretion would be reduced because officers might be expected to stop every driver they see committing a traffic violation. One could also imagine low-tech changes with the potential to reduce officer discretion. Perhaps we could ex ante require officers to only stop every fifteenth driver they notice commit a low-level

\begin{footnotes}

\textsuperscript{141} See id. at 1260 (“[S]upervisors are reluctant to file complaints against subordinates, understanding that doing so puts their subordinates’ future chances of promotion at risk and likely reflects poorly on the supervisor.” (citation omitted)).

\textsuperscript{142} For a discussion of pros and cons of civilian review boards, see Udi Ofer, \textit{Getting It Right: Building Effective Civilian Review Boards to Oversee Police}, 46 SETON HALL L. REV. 1033, 1038 (2016).

\textsuperscript{143} See Harris & Worden, supra note 140, at 1280–81.


\end{footnotes}
traffic violation. Of course, such a solution would still require significant trust in the police. But it might be a step in the right direction.

We should note, this is not an endorsement of any of these solutions. Our primary goal in this Article has been to assess the conventional wisdom that Fourth Amendment-like standards have a meaningful bearing on officer behavior. Having cast doubt on that prediction, we simply seek to offer certain alternative approaches that do not suffer from the setbacks of standards enforced via the exclusionary rule. Ultimately, approaches like those suggested should be the subject of future inquiry and research.

**CONCLUSION**

A dominant view among scholars of the Fourth Amendment is that a permissive stance towards pretextual traffic stops contributes to greater disparities in the rates of traffic stops across race. We suggest, however, that there may be conceptual grounds to doubt this conventional view and stress the weaknesses in the deterrent channel linking the substantive constitutional law and the officer decision-making process. Ultimately, whether or not the imposition of a Fourth Amendment-like prohibition against pretextual stops will reduce the racial disparities observed in such stops is an empirical question. Drawing on a natural experiment based on a change in the constitutional treatment of pretextual stops in the State of Washington and employing a range of empirical techniques, we fail to find evidence in support of the conventional wisdom. To the extent that policymakers wish to address the undeniable problem of racial bias in the rates of traffic stops, our evidence cautions against placing all of our hopes in a constitutional-rights-based approach. This is not to say that such rights are unimportant. They simply are not enough to attack this critical issue. We propose that policymakers consider certain extraconstitutional approaches, including those better designed to bolster a deterrent channel or those that attempt to remove officer discretion from traffic regulation altogether.