MPC

Minnesota Population Center

UNIVERSITY OF MINNESOTA

Racial Bias in Traffic Stops: Tests of a Unified Model of Stops and Searches

Joseph A. Ritter
Department of Applied Economics
University of Minnesota

June 2013

Working Paper No. 2013-05
https://doi.org/10.18128/MPC2013-05

I am grateful for constructive comments from Tim Beatty, Juan Chapaorro, Deborah Levison, Paul Glewwe, Sam Myers, and Aaron Sojourner. The Center for Urban and Regional Affairs at the University of Minnesota provided financial support for part of this research.

Abstract

This paper develops a model of traffic stops and subsequent searches in which police officers use information about the race of drivers to maximize a well defined objective. The model provides a behavioral foundation absent from Grogger and Ridgeway's (2006) elegantly simple test and, by incorporating searches, adds two complementary tests. Using data collected during 2002 by the Minneapolis Police Department, the tests rule out (1) statistical discrimination, (2) taste-based discrimination by optimizing police officers, and (3) statistical discrimination with cognitive limitations. The pattern of results is consistent with implicit discrimination.

The question of racial bias in law enforcement is a vexing social, policy, and statistical problem. Throughout the country, police departments have introduced policies intended to deter racial profiling, yet the phrase "driving while black" is still often heard, and high-profile incidents occasionally reach the headlines. Does race, in itself, affect police decisions? If so, what motivates differential treatment—statistical discrimination or something else, such as some form of racial prejudice?[1]

There are, unquestionably, racial disparities in traffic stops. However, attributing these disparities to race per se—establishing a racial bias—is difficult. The central statistical problem is that most of the confounding variables are not observed; an important reason why debates between activists and police continue is the difficulty in achieving convincing statistical identification, both for detecting racial bias in stop rates and for distinguishing among possible reasons for a bias. Grogger and Ridgeway (GR, 2006) developed an elegantly simple test for racial bias in stop rates that bases identification on the fact that darkness impedes recognition of race. By restricting the sample to stops that occur between the times of the earliest and latest sunset of the year, the method creates a natural experiment (section 1.1).

The present paper develops a theoretical model of the use of race in stop and search decisions. Building on GR's natural experiment, the model emphasizes the role of light and darkness in concealing and revealing drivers' race in the context of an officer's choice problem: race is sometimes observed before a vehicle is stopped, but after the stop takes place the police officer is always able to form an definite opinion about the driver's race.[2] Thus, on average, a stop reveals more information about the driver's race or ethnicity during darkness than during daylight. Two new empirical tests follow from this observation. The combination of the new tests with GR's test allows deeper insight into the nature of possible racial bias.

In economics, most previous research on racial bias in traffic stops falls into

[1] The term "racial profiling" has been used both to denote *any* use of race by police and in the narrower sense of statistical discrimination used as a law-enforcement tool. I therefore avoid it except in referring to sources which themselves use the term.

[2] To the extent that the driver's racial self-identification differs from the officer's opinion, it is the latter that is relevant here.

two strands. One strand leverages theoretical models of police decision making to derive empirical predictions of statistical discrimination based on relationships among search rates and/or rates of finding contraband in searches of different demographic groups. The hit-rate test advanced by Knowles, Persico and Todd (KPT, 2001) is the seminal example of this line of analysis. KPT's theoretical model implies that the rate at which contraband is found in searches (the hit rate) should be the same for different demographic groups, if police use statistical discrimination as a law-enforcement tool.

The second strand, exemplified by Antonovics and Knight (2009), bases identification of prejudice on the principle that search decisions based on statistical discrimination should be independent of officer race; they find that mismatches between driver's and officer's race affect the probability of search. Close and Mason (2006) also find that mismatches between officer's and driver's race have a significant impact on enforcement actions that result from a stop. Anwar and Fang (2006) use officers' race, but take a nonparametric approach, as described in section 1.2.

An important aspect of most of the papers cited in the previous two paragraphs is that they divide reasons for racial disparity into statistical discrimination and "prejudice." Strictly speaking, however, the alternative hypothesis is that optimal statistical discrimination is not used. This coarse partition of possibilities hides subtle distinctions. Statistical discrimination can be imperfectly implemented; prejudice can be limited in scope; cognitive biases can be triggered.

By using information about both stops and searches, the combination of tests I use can and, in fact, does make finer distinctions. Applied to data collected during 2002 by the Minneapolis Police Department, the tests provide strong evidence against (1) statistical discrimination, (2) taste-based discrimination by optimizing police officers (that is, a racial preference on which officers act consistently), and (3) statistical discrimination with cognitive limitations (in a sense explained below).[3] The pattern of results is consistent with implicit dis-

[3]The traffic stop data used in this paper were collected by the Minneapolis Police Department during 2002 because of contemporary concern about racial profiling in Minnesota. The Department subsequently instituted efforts to avoid profiling, some of which were outlined in a federal mediation agreement in December 2003. There was also a change in Department leadership at the beginning of 2004. More recent data would be required to evaluate the

crimination (Bertrand, Chugh, and Mullainathan, 2005), but may also fit other explanations not considered here.

The paper proceeds as follows. Because GR did not base their test on an explicit model of police decision making, the first (stop) stage of the model developed in section 2 elucidates conditions sufficient to justify GR's test for racial bias in traffic stops, while the second (search) stage derives additional testable implications for how search rates should change between daylight and darkness. The data are described in section 3. Section 4 tests the model's implications for stops and searches and reports sensitivity analyses. Section 5 discusses implications of the empirical results.

1 Previous research

This section discusses in more detail the two strands of prior research discussed above.

1.1 The "veil-of-darkness" methodology

Grogger and Ridgeway's method for detecting racial bias starts with the commonsense observation that recognizing the race of a driver is more difficult in darkness than in daylight. A simple day-night comparison is insufficient, however, because driving and policing patterns differ throughout the day. There is no reason to think these variations are race-neutral and they may interact. Moreover, data on these patterns are not generally available.

GR's refinement of the day-night comparison creates a natural experiment. They noted that there are times of day that are before sunset during the summer and after sunset during the winter (also between earliest and latest sunrise, but there are many fewer traffic stops early in the morning). Thus they propose discarding stops that did not happen between the time of earliest sunset and

impacts of these changes.

latest sunset; they refer to this as the "intertwilight period." In Minneapolis the evening intertwilight period is roughly between 5 and 9 P.M. Within the intertwilight period, if there is a bias against, say, black drivers, the fraction of drivers stopped who are black should be lower during darkness. The underlying logic is that of difference-in-differences with darkness as the treatment. In other words, the empirical question is: How does the the experience of nonwhite drivers change relative to white drivers when it is dark?

GR argue in their paper that the veil-of-darkness methodology cuts through most of the statistical issues surrounding the detection of racial bias in traffic stops, but they do not offer a theoretical model that elucidates the behavioral assumptions necessary to justify this claim. Their primary identifying assumption is that there are not significant seasonal changes in driving and policing during the intertwilight period. That means, specifically, that changing light levels do not significantly change policing patterns or the population of drivers on the road or differentially alter driving behavior. Functions of the date of the traffic stops can be used to check for significant seasonal changes. Although lack of seasonal changes is the key assumption, others not explicitly specified by GR are needed. In section 2.2, I describe a model that leads to GR's hypothesis. This model is also the first stage of the model of searches that I develop in section 2.3.

In the Minneapolis data, officers reported that they could identify the race of the driver only 19.4 percent of the time during daylight and only 9.4 percent of the time during darkness.[4] The fact that these rates are so low leads to the main weakness of GR's method: the test can have low power because an observation carries no information if the officer could not identify the race of the driver ex ante. Also, as GR point out, if police are able to infer the race of the driver from characteristics of the vehicle that are visible in darkness, the power of the test will be lower (because the race recognition gap between daylight and darkness is reduced). GR's test has been applied to data from Oakland, California (Grogger and Ridgeway, 2006) and to Cincinnati, Ohio (Schell, et al., 2007). No evidence of racial bias was found in either case. Ritter and Bael (2009) applied the test to the Minneapolis data used here and found strong evidence of racial bias.

[4]Although the data on whether race was known could have been misreported, personal experimentation suggests that these figures are in the right ballpark. It is often impossible to identify the race of the driver of another car during the day because of position, glare, tinting, and so forth. Similarly, it is occasionally possible to see the race of a driver at night.

An important limitation of GR's "veil of darkness" test is that it can establish the existence of racial bias in stops, but offers no information about reasons for the bias.

1.2 Searches and prejudice

Knowles, Persico and Todd (2001) developed a model of optimal policing, which implies equalization across demographic groups of the success rates of searches ("hit rates"). Departures from equality are interpreted as evidence of prejudice. KPT found that data on stops by the Maryland State Police were "consistent with maximizing behavior by police who are not racially prejudiced." Here I highlight several assumptions behind the hit-rate test that can be relaxed using the empirical strategy proposed in this paper.

The logic of hit-rate tests starts with the premise that in the absence of prejudice the cost of stopping and searching a motorist is the same across demographic groups. The cost of searching represents a combination of psychological costs or benefits and the opportunity cost of the time spent on the search (for example, the delay in returning to traffic enforcement). If the cost of searching is equal across groups, it is optimal to adjust search rates to equalize the marginal probability of a successful search. The subtle insight of KPT's paper is that in their model the response of drivers to optimal policing implies equalization of the average probability, not just the marginal probability, a fact that is extremely valuable for empirical testing.

The implication of equalized hit rates requires several additional assumptions. First, it is actually necessary that the expected *net benefits* of searches are the same for different demographic groups in the absence of prejudice. Thus KPT also assume that the expected (gross) payoff from a successful search is the same among demographic groups. But it seems unlikely, for example, that police view payoffs from finding a large shipment of cocaine and finding a single marijuana seed as equal, and it is not implausible that the distribution of crime severity differs among demographic groups.[5] However, deviations from equal hit-rates

[5]Barbe and Horrace (2012) discuss this issue in their critique of Knowles et al., though the

have been interpreted as evidence that the *cost* of searching differs—that there is racial prejudice in the decision to search.

Second, the test assumes that search costs are constant except for the possibility of a taste for discrimination. There are two reasons this assumption may be too strong. First, it is likely that the opportunity cost of searches varies during the day, and this variation could interact with racial differences in driving patterns (one of the many confounding variables alluded to above). For example, the cost of a search is probably different at midnight than during rush hour, as is the racial composition of drivers on the road. This kind of interaction would undermine the validity of a hit-rate test that aggregates hit rates over wide time intervals.

A second reason to question the assumption that search costs are constant is that spatial variation in policing often reflects intentional policy and may be correlated with race in ways that do not imply prejudice. For example, in 1998 the Minneapolis Police Department introduced a crime-reduction strategy called CODEFOR, which uses the spatial distribution of 911 calls and crime reports to allocate police resources (Myers, 2002). Consequently, several areas subsequently received high-intensity law enforcement, and the residents of most of these areas were disproportionately racial minorities. More police patrols increased the number of traffic stops, and the drivers were more likely to be people of color relative to the population of Minneapolis. The intense focus on localized crime reduction may have also lowered the threshold for searches, possibly lowering the city-wide hit rate for racial minorities.

Finally, and perhaps most important, is a limitation that KPT themselves noted: The model assumes that motorists respond optimally to the probability of being searched, given their characteristics. To do so would require a great deal of information—much more than is available for the present study, for example—and a high level of rationality. KPT point out that without full response by motorists to search intensity, the model only implies equalization of marginal probabilities, making it empirically intractable.

Anwar and Fang (AF, 2006) question the basic hit-rate approach from a differ-

connection to KPT's model is not explicit.

ent direction. They point out that many data sources, including the Maryland State Police data used by KPT, do not identify the race of the officer. This forces the test to implicitly assume that search costs and interpretation of of law-enforcement information do not vary by race of the police officer. Using data gathered by the Florida Highway Patrol, they find evidence against "monolithic behavior" of officers, i.e., evidence that officers of different races do not face the same costs.

AF then used police officers' race as a means to relax some of the assumptions needed for hit-rate tests. Their key insight was that, while officers of different races might interpret information differently and may face different conditions, the ordering of search rates or hit rates across *officers* of different races should be the same, regardless of whether they are considering white, black, or Hispanic drivers. This is an important, but subtle, innovation because by using a non-parametric approach they avoid the infra-marginality problem that necessitates the strong assumptions of KPT's model.

AF's result seems counter-intuitive because it compares officers, rather than motorists, but the intuition is relatively straightforward when approached from the right angle. Let $t(r_m, r_p)$ be the cost for an officer of race r_p of searching a motorist of race r_m. Following KPT, AF define absence of prejudice by $t(M, r_p) = t(W, r_p)$ for minority (M) and white (W) drivers. Unlike KPT, however, they do not assume that $t(r_m, M) = t(r_m, W)$; costs can differ between officers of different races.

Suppose for illustration that officers of neither race are prejudiced, but that W officers face higher search costs for some reason. Thus

$$t(M, M) = t(W, M) < t(M, W) = t(W, W) \tag{1}$$

(the equalities indicate lack of prejudice). In AF's model, officers receive a signal about "guilt," and it follows from (1) that W officers will set a laxer threshold than M officers for searches of *both races*. The laxer threshold translates into a higher hit rate for *both races* of driver. Violations of these orderings is taken as evidence that at least one of the equalities in (1) is violated, i.e., of prejudice, but does not determine who is prejudiced against whom.[6] Also, as AF point out,

[6]Although both papers use the race of the officer to help identify prejudice, this logic is

the orderings could be consistent in the presence of any level of prejudice that does not flip the inequalities $t(M, W) > t(M, M)$ and $t(W, W) > t(W, M)$ (thus their tests can have low power).

That weakness is also a strength, however, because for the same mathematical reason, AF's tests are robust to the possibility that search costs differ systematically by race of driver. To see this, suppose that

$$t(M, M) = t(W, M) + a < t(M, W) = t(W, W) + a,$$

where a is a constant. This variation maintains the inequalities $t(M, W) > t(M, M)$ and $t(W, W) > t(W, M)$, so W officers would still use a laxer standard against motorists of both races.

AF's method has the additional advantage of not presupposing that criminally inclined motorists act optimally based on extensive information about police practices. However, it has two disadvantages in addition to low power: Their method does not address the problems of spatial and temporal heterogeneity discussed above. Finally, in the absence of prejudice, AF offer no way to determine whether statistical discrimination is in use.

2 A solar-powered model of stops and searches

If police officers systematically use race either for statistical discrimination or to exercise a taste for discrimination, they have two opportunities to do so: in deciding whether to stop a vehicle and in deciding whether to initiate a search. This section develops an integrated model of these decisions, which emphasizes how light alters the flow of information about race. The model leads to GR's hypothesis about stops and two additional hypotheses about searches.

The model presented below focuses on optimal response to an officer's recognition of a driver's race. Therefore, although I mainly present the model as one of

distinctly different than that of Antonovics and Knight (2009). Antonovics and Knight assume that prejudice takes the form $t(M, M) > t(W, M)$ and $t(W, W) > t(M, W)$.

statistical discrimination, two key assumptions can be also be interpreted as describing a taste for discrimination. Consequently, the model does not distinguish between statistical discrimination and optimally exercising a taste for discrimination. The two are observationally equivalent because I do not employ "guilt" outcomes, such as hit rates.

Apart from GR, the null hypothesis in the papers described above is a relationship implied by statistical discrimination. The alternative hypothesis is "not statistical discrimination," but has typically been labeled "prejudice." In the present paper the null is wider: that police officers use the driver's race to maximize a well defined objective. The different alternative hypothesis highlights that prejudice does not necessarily result in behavior that can be described as rational (in the economist's sense) and that attempts at optimal law enforcement may fall short due to cognitive or other limitations.

The overall framework of the model is as follows. A police officer first observes some characteristics and behaviors of a driver and—sometimes—race. The officer then decides whether or not to stop the vehicle. If a stop takes place, the officer observes the driver's race as well as additional characteristics and behaviors, then decides whether to search the driver and vehicle.

2.1 Preliminary example

To set the intuition behind the tests, consider what happens in a short interval during the intertwilight period, say between 6:00 and 6:15 P.M. To make this example as transparent as possible, I temporarily make the following extreme assumptions:

1. Considering only this interval acts as a perfect natural experiment—only the amount of light changes between daylight and darkness. Race is perfectly observable when it is light, but never observable when it is dark.

2. Police officers employ statistical discrimination, based on actual differences in the propensity for drivers of different races to be carrying contraband.

They value all successful searches equally. When deciding whether to initiate a stop, they mentally combine all observed characteristics and behaviors other than race into a single index J that orders the predicted probability a driver is carrying contraband.

3. Conditional on all other characteristics, group n (nonwhite) has a higher propensity to carry contraband than group w (white).

4. There is a cost to stopping a motorist and an additional cost to search the vehicle. Traffic enforcement is not a concern.

5. After a stop takes place, if it is not already known, the race of the driver is revealed and is the only relevant information revealed.

During daylight, officers will set race-specific thresholds J_n and J_w to determine whom to stop and search. These thresholds will be set so that, at the margin, the probability of finding contraband is the same for both races:

$$\Pr(\text{contraband}|J_n, n) = \Pr(\text{contraband}|J_w, w).$$

Since knowing a driver is of race n raises the conditional probability he will have contraband, the threshold for n drivers is lower than that for w drivers: $J_n < J_w$.

During darkness, race is unknown prior to the stop, so officers set a common intermediate threshold for both races, J_u. Since $J_n < J_u < J_w$, the fraction of stopped drivers who are race w is higher in darkness. This is what the GR's methodology looks for.

After the stop, however, the race of the driver is revealed. Some of the w drivers have $J < J_w$ and, in order to avoid the search cost, are not searched. Thus during the 6:00–6:15 interval,

$$\Pr(\text{search}|\text{stop}, w, \text{daylight}) > \Pr(\text{search}|\text{stop}, w, \text{darkness}). \tag{2}$$

This implication is driven by the fact that race is revealed by the stop. In this minimal model, all n drivers who are stopped are also searched, so there is not a parallel inequality, but in the more general model of section 2.3,

$$\Pr(\text{search}|\text{stop}, n, \text{daylight}) < \Pr(\text{search}|\text{stop}, n, \text{darkness}). \tag{3}$$

Thus statistical discrimination implies both that the proportion of stopped drivers who are n will be higher during daylight than darkness and also inequalities (2) and (3). I now turn to a more formal derivation of the testable hypotheses.

2.2 Decision to stop

This section presents a model of the decision to stop a motorist, which leads to GR's veil-of-darkness test. I begin with assumptions relevant to the decision to stop a driver. Most have analogues in the search stage, but I defer stating those until section 2.3.

Prior to a stop, the officer observes characteristics and behaviors x (shortened to "characteristics" in the sequel), not including race. The officer also observes a random variable, $R \in \{w, n, u\}$, that is an indicator of whether the driver's race is white (w), nonwhite (n), or unknown (u). For simplicity I assume that officers do not make errors (e.g., identifying a w driver as n) or probabilistic judgements of race (for example, "The probability that driver is white is 0.7.").[7]

Let $\ell = 1$ indicate daylight and $\ell = 0$ darkness. Race is observed with probability δ during daylight and η during darkness, i.e., $\Pr(R = u|\ell = 1) = 1 - \delta$ and $\Pr(R = u|\ell = 0) = 1 - \eta$. Until stated otherwise, time of day (t) is held constant; I suppress this time notation until needed.

The police receive benefits from traffic enforcement and from criminal law enforcement, but the process of stopping and searching a motorist has psychological costs, potential legal costs, and opportunity costs (such as foregone traffic

[7]The statement that officers do not make probabilistic judgements refers to direct observation of race. The vector x may contain information correlated with race and, in any case, police officers would have priors based on population proportions.

enforcement). Let A be the expected traffic enforcement benefit for a potential stop. Let $B(x, R)$ represent all other expected benefits and costs of making the stop, including the expected cost of the stop, the cost of a possible search, and the possible benefit of a successful search (discovering contraband, for example). Although the distinction between law-enforcement and/or personal benefits of the stop (B) and traffic-enforcement benefits (A) does not play a prominent role in the analysis, it is important because the latter do not play a role in search decisions. It is not necessary to assume that either A or B (or $A + B$) is independent of race, time of day, or location as assumed (explicitly or implicitly) by many previous papers.

Assumption 1. *Officers order characteristics such that if x ranks higher than y, $B(x, R) > B(y, R)$ for any value of R. The ordering meets standard conditions on preference orderings: completeness, transitivity, and reflexivity.*

This assumption means, first, that officers consistently evaluate the relationship between non-race characteristics and the payoff from search and, second, that these characteristics affect the perceived payoff in the same direction, regardless of race.[8] Although the assumption says the ordering of characteristics is the same for different races, it does not say that the distribution of characteristics is invariant to race.

For notational convenience, represent the ordering of characteristics by a scalar function $J(x)$ and henceforth write $B(J, R)$ rather than $B(x, R)$. From the way J is defined, $B(J, R)$ must be increasing in J.

Assumption 2. *Officers stop a vehicle when $A + B(J, R) > 0$ and there is some x in the population of drivers for which $B(J(x), R) > 0$.*

The latter part of of Assumption 2 is technically needed to ensure that B is not bounded below zero, but is mild; it simply formalizes the idea that police officers might observe very bad behavior.

[8]Anwar and Fang (2006) instead assume the non-race information received by police officers satisfies the monotone likelihood ratio property. They state that the assumption is without loss of generality because the scalar index can always be reordered to make it satisfy the MLRP. However, since it may not be possible to achieve that goal using the same reordering for both races, I directly assume the required consistency in the ordering.

Assumption 3. *Race changes the payoff of stops:* $B(J_1, n) = B(J_2, w)$ *implies* $J_1 < J_2$.

Assumption 3 can be interpreted as saying either that officers believe that race is informative about law-enforcement or that police officers have a taste for discrimination. I now turn to the key identification assumptions.

Assumption 4. *Darkness impedes recognition of race:* $\eta < \delta$.

Assumption 5. *The intertwilight period acts as a natural experiment. Specifically, during the intertwilight period, the following do not depend directly on whether it is dark or light: (i) the rate at which police observe* n *motorists,* $N(t)$*; (ii) the rate at which police observe* w *motorists,* $W(t)$*; (iii)* $B(J, R, t)$*; and (iv) the joint distribution of* J *and* A*, conditional on* R*,* $F(J, A|R)$.[9]

GR's innovation that the intertwilight period can be treated as a natural experiment is embedded in this assumption. In statistical language it says, among other things, that comparing outcomes between daylight and darkness adequately controls for the mix of drivers on the road, their vehicles, and their behavior. Similarly, policing might be more intensive in certain neighborhoods or certain times of day, but this assumption does not require absence of this geographic or intra-day variation, only that it is stable. For example, this assumption is not invalidated by the existence of the CODEFOR program mentioned above; it only requires that the program did not appear or disappear during the period under study. I address possible failures of this assumption in several ways in the empirical analysis.

Part (iv) is the strongest part of the assumption, saying essentially that, prior to the stop, relevant information other than race is equally observable during darkness and daylight. In other words, the only way darkness affects the probability of stops is the relative observability of race.

When race is not observed, the expected net benefit of a stop is given by

[9] As mentioned above, the absence of dependence darkness and light is really about absence of systematic seasonal changes but it is easier to frame the theory in this binary way. The possibility of residual seasonality is addressed empirically in section 4.

$A + B(J, u) = A + bB(J, n) + (1 - b)B(J, w)$ with $b \in (0, 1)$. The probability weight b depends on the proportions of drivers on the road, which may vary with t, but I assume for simplicity that it does not depend on J. In principle, Bayesian police officers could calculate $\Pr(r|J)$, but that would add needless complication to the model since the key thing for the results below is that $0 < b < 1$.

$B(J, R)$ is increasing in J and not bounded below zero (Assumption 2). Therefore, given Assumptions 3 and 5, police officers' decision rules specify thresholds $J_n(A) < J_u(A) < J_w(A)$ defined by $A + B(J_r(A), r) = 0$, such that drivers observed to be n are stopped if $J > J_n(A)$, drivers observed to be w are stopped if $J > J_w(A)$ and drivers whose race is not observed are stopped if $J > J_u(A)$. In the absence of part (iv) of Assumption 5, the thresholds would depend on ℓ as well as R and A.

Several conditional probabilities are used in the remainder of this section. The notation is defined as follows:

- $\Pr(\text{stop}|r, r)$ is the probability an r driver ($r \in \{n, w\}$), recognized as r, is stopped. Assumption 5 implies that this probability does not depend on ℓ (recall that time of day is held constant).

- $\Pr(\text{stop}|u)$ is the probability that a driver of unknown race is stopped. Since it is a weighted average of $\Pr(\text{stop}|n, n)$ and $\Pr(\text{stop}|w, w)$ (specifically, $\Pr(\text{stop}|u) = b \Pr(\text{stop}|n, n) + (1 - b) \Pr(\text{stop}|w, w)$), the probability does not depend on ℓ.

- $\Pr(\text{stop}|r, \ell = 1) = \delta \Pr(\text{stop}|r, r) + (1 - \delta) \Pr(\text{stop}|u)$ is the probability an r driver is stopped in daylight ($\ell = 1$), not conditioned on whether race is identified. Also, $\Pr(\text{stop}|r, \ell = 0) = \eta \Pr(\text{stop}|r, r) + (1 - \eta) \Pr(\text{stop}|u)$.

Given that officers stop drivers when $A + B(J, R) > 0$ and that $B(J, R)$ is increasing in J,

$$\Pr(\text{stop}|n, n) = \int_{\forall A} \int_{J=-\infty}^{\infty} 1[A + B(J, n) > 0] dF(J, A|n) = \int_{\forall A} \int_{J=J_n(A)}^{\infty} dF(J, A|n)$$

and

$$\Pr(\text{stop}|w, w) = \int_{\forall A} \int_{J=J_w(A)}^{\infty} dF(J, A|w).$$

I now turn to implications of the model. Proposition 1 is a standard implication of statistical discrimination, which I state without proof.[10]

Proposition 1. *Holding time of day constant,*

$$\Pr(\text{stop}|n, \ell) > \Pr(\text{stop}|w, \ell).$$

Although Proposition 1 is an intuitive result, it is too weak to be useful and illustrates the core statistical problem in detecting racial bias in stops. Even if the inequality could be empirically tested (it is difficult to measure the denominators), the inequality could hold, even if police officers do not use race in their decisions, because the distribution of J differs between n and w drivers. GR's natural experiment addresses this crucial identification question, and I now derive the empirical implication of the model that leads to GR's test.

Proposition 2. *Holding time of day constant,*

$$\Pr(\text{stop}|n, \ell = 1) > \Pr(\text{stop}|n, \ell = 0)$$
$$\Pr(\text{stop}|w, \ell = 1) < \Pr(\text{stop}|w, \ell = 0).$$

Proof. Since $J_n(A) < J_u(A) < J_w(A)$ for all A, $\Pr(\text{stop}|n, n) > \Pr(\text{stop}|u) > \Pr(\text{stop}|w, w)$ and, therefore,

$$\begin{aligned}\Pr(\text{stop}|n, \ell = 1) &= \delta \Pr(\text{stop}|n, n) + (1 - \delta) \Pr(\text{stop}|u) \\ &> \eta \Pr(\text{stop}|n, n) + (1 - \eta) \Pr(\text{stop}|u) \\ &= \Pr(\text{stop}|n, \ell = 0)\end{aligned}$$

[10] This proposition, which I use only for illustration, actually requires one more assumption (not required for the remainder of the paper) to rule out the possibility that racial differences in the distribution of J do not offset the effects of Assumption 3. That scenario would be the reverse of what is claimed by police departments with policies prohibiting profiling—that large racial disparities in stops occur mainly because nonwhite drivers are more likely than white drivers to behave in ways that induce stops.

and[11]

$$\begin{aligned}\Pr(\text{stop}|w,\ell=1) &= \delta\Pr(\text{stop}|w,w)+(1-\delta)\Pr(\text{stop}|u)\\ &< \eta\Pr(\text{stop}|w,w)+(1-\eta)\Pr(\text{stop}|u)\\ &= \Pr(\text{stop}|w,\ell=0).\end{aligned}$$

This result expresses the key intuition behind GR's test: if police officers use race in stop decisions, stops of n drivers are relatively more frequent during daylight. As mentioned above, this result uses difference-in-differences logic. Rather than comparing what happens to n and w drivers at a given time of day as does Proposition 1, it compares how things *change* for n and w drivers as a treatment (darkness) is applied.

Proposition 2 considers only what happens at a particular time of day during the intertwilight period. Proposition 3 completes the picture by aggregating over an interval $T=[t_1,t_2]$ that is part of the intertwilight period.

PROPOSITION 3. *The proportion of drivers stopped during T who are n is higher when $\ell=1$ than when $\ell=0$.*

Proof. For clarity I do not suppress the dependence of probabilities on time of day, t. The proposition is equivalent to

$$\frac{\int_{t\in T}\Pr(\text{stop}|n,\ell=1,t)N(t)\,dt}{\int_{t\in T}\Pr(\text{stop}|w,\ell=1,t)W(t)\,dt} > \frac{\int_{t\in T}\Pr(\text{stop}|n,\ell=0,t)N(t)\,dt}{\int_{t\in T}\Pr(\text{stop}|w,\ell=0,t)W(t)\,dt}.$$

Assumption 5 says that $N(t)$ and $W(t)$ (the rates at which n and w drivers are observed by police) do not depend on ℓ. Proposition 2 says that the numerator on the left (which is the number of n drivers stopped during daylight) is pointwise

[11] Part (iv) of Assumption 5 simplifies justification of the inequality to a comparison of δ and η. Without the assumption, $\Pr(\text{stop}|n,n)$ and $\Pr(\text{stop}|w,w)$ would both depend on ℓ, and without some restriction about the joint distribution of ℓ and J, it might be possible that change in the distribution of J between $\ell=0$ and $\ell=1$ fully offsets the effect of ℓ on observability of race. This seems unlikely, but is obviously not testable.

greater than the numerator on the right and that the denominator on the left is pointwise less than the denominator on the right, which verifies the inequality.

2.3 Decision to search

In stopping a vehicle, the officer has in essence purchased an option on a search or some other law enforcement action. Once a stop has taken place, the officer observes additional characteristics and behavior, and the driver's race is always observed. The officer then chooses whether to exercise the option.[12] The structure of the decision to search is very similar to the decision to stop a vehicle, except that race is always known, and traffic-enforcement considerations are no longer relevant. A key difference, however, is that the population of drivers about which search decisions are made is the population of stopped drivers rather than all drivers. The fact that the population of stopped drivers is a selected sample is key to the model.

Denote the combined characteristics observed before and after the stop by x^* and the post-stop observation of race by $R^* \in \{n, w\}$. Any response of the driver to being stopped is incorporated in x^*. Let $B^*(x^*, R^*)$ be the expected net benefit of a search, given information x^* and observed race R^*. Again, it is not necessary to assume that the net benefit is independent of race, time of day, or location. The first few assumptions below parallel earlier assumptions about stops.

Assumption $1'$. *Officers order post-stop characteristics such that if x^* ranks higher than y^*, $B^*(x^*, R^*) > B^*(y^*, R^*)$ for any value of R^*. The ordering meets standard conditions on preference orderings: completeness, transitivity, and reflexivity.*

Following the previous section, represent the ordering by $J^*(x^*)$ and write $B^*(J^*, R^*)$ rather than $B^*(x^*, R^*)$.

Assumption $2'$. *Officers search a vehicle when $B^*(J^*, R^*) > 0$ and there is*

[12] The option language suggests the possibility of statistical discrimination based on the variance B being higher for n drivers than w drivers, but I do not pursue that angle here.

some x^* for which $B(J^*(x^*), R) > 0$.

Assumption 3′. *Race is useful in searches: $B^*(J_1^*, n) = B^*(J_2^*, w)$ implies $J_1^* < J_2^*$.*

Like Assumption 3, this is also a way of specifying a taste for discrimination.

Assumption 5′. *Within the intertwilight period, the following do not depend directly on whether it is dark or light: (i) $B^*(J^*, R^*, t)$; and (ii) the distribution of J^* conditional on R^*, $F^*(J^*|R^*)$.*

The language of part (ii) of Assumption 5′ is identical to part (iv) of Assumption 5, but this assumption is much weaker in practical terms since the officer can inspect the vehicle and driver at close range after the stop and can use artificial illumination. The final assumption says that bad news before the stop raises the likelihood of bad news after the stop:

Assumption 6. *If $J_2 > J_1$, then $J^*|J_2$ has first-order stochastic dominance over $J^*|J_1$.*

Assumption 6 is really an assumption that the orderings J and J^* are formed in a reasonable way, rather than an assumption about the statistical relationship between x and x^*.

$B^*(J^*, R^*)$ is increasing in J^* and, therefore, Assumption 3′ implies that there are two thresholds, $J_n^* < J_w^*$, defined by $B^*(J_r^*, r) = 0$, such that an n driver is searched if $J^* > J_n^*$ and a w driver is searched if $J^* > J_w^*$. Assumptions 1′ through 5′ and 6 imply the following testable implications of the use of race in search decisions.

Proposition 4. *Let T be a time interval during the intertwilight period. Then during T:*

(i) A higher fraction of stopped n drivers are searched during darkness than during daylight:

$$\Pr(\text{search}|\text{stopped}, \text{n}, \ell = 0, \text{T}) > \Pr(\text{search}|\text{stopped}, \text{n}, \ell = 1, \text{T}).$$

(ii) A lower fraction of stopped w drivers are searched during darkness than during daylight:

$$\Pr(\text{search}|\text{stopped}, \text{w}, \ell = 0, \text{T}) < \Pr(\text{search}|\text{stopped}, \text{w}, \ell = 1, \text{T}).$$

The proof is long and can be found in the appendix, but the intuition behind the proposition can be stated relatively simply. The population of stopped drivers is a censored sample. For n drivers it consists of drivers with $J > J_n(A)$ or $J > J_u(A)$, according to whether their race was observed or not prior to the stop. Darkness changes the censoring and shifts the distribution of J in that sample toward the J_u threshold (because the race of the driver is less visible during darkness). This means that during darkness stopped n drivers have higher average J than during daylight. Assumption 6 implies they have higher average J^* as well, so a higher percentage are searched during darkness.

Parallel logic for w drivers works in the opposite direction. Darkness means more w drivers with $J \in \big(J_u(A), J_w(a)\big)$ are stopped, so the population of stopped w drivers has lower average J and, thus, lower average J^* than during daylight, so fewer searches are conducted during darkness.

It may seem surprising that Bayes' Theorem was not invoked here since the model of the search decision clearly describes a process of Bayesian updating. In fact, Proposition 4 is a simple aggregate implication of Bayesian updating by police officers. But, as the previous paragraphs should make clear, it is really a property of a decision process that uses race to maximize—even only approximately—a well defined objective function. Police officers do not have to derive the optimal thresholds for stops and searches. They only need to get them in the right order ($J_n(A) < J_u(A) < J_w(A)$ and $J_n^* < J_w^*$) for Propositions 3 and 4 to hold. I discuss the implications of that observation after reporting the results of the tests.

3 Data

As part of a state-sponsored study, the Minneapolis Police Department gathered data on every traffic stop that took place during 2002, a total of 53,559 stops.[13] Although data were collected in other jurisdictions, I use only Minneapolis to avoid complications arising from mixing jurisdictions. Also, Minneapolis recorded the most stops of any jurisdiction by a factor of four (the St. Paul Police Department did not participate). Each record includes the race of the driver as perceived by the police officer after the stop was completed (i.e., after direct observation of the driver), the time and date, the reason for the stop, and, if a search took place, details about the reason for and outcome of the search. Officer race was not recorded.

All of the data are from reports by the police officer who conducted the stop, but these data were not integrated into the normal record-keeping process of the state or city so there is no way to check the accuracy of the data. Nor did police cars have video cameras at the time. (In fact cameras were purchased in 2004 with a grant awarded for participation in the racial profiling study.) Given the conflicting demands on officers, it seems likely that some stops were not recorded, but there is no reason to think that this created any statistical bias.

In the data, race of the driver is based the officer's perception, not self-classification by the driver. Drivers' self-classifications are not available in the data, but officers' classifications would be preferred here in any case since officers are the decision makers whose choices are being examined for bias. A separate issue is whether officers accurately recorded the race they believed the driver to be, but there is no evidence of any falsification.

As illustrated by table 1, racial differences in the raw data are stark: 39.7 percent of drivers stopped were black, but only 15.8 percent of the driving-age population in 2000 was black; 44.7 percent of drivers stopped were white, but 69.8 percent of the driving-age population was white. Discretionary searches, defined in two ways, are even more concentrated: fully 80 percent of these searches were of nonwhite drivers.

[13] See Myers (2002) for a discussion of the context in which the data were collected.

Table 1: Stops and searches by race
(percent of total)

	Black	Hispanic	White	Nonwhite[a]
Driving age population[b]	15.8	6.6	69.8	30.2
Stops (all hours)	39.7	10.7	44.7	55.3
Stops during intertwilight interval				
Stops	39.2	12.7	42.8	57.2
Discretionary searches				
Narrow definition[c]	61.1	15.3	19.1	80.9
Wide definition[c]	57.3	19.8	18.6	81.4

[a]Nonwhite includes Asian and Native American drivers.

[b]From CCJ/IRP (2003), based on 2000 Census.

[c]For definition see section 3.2.

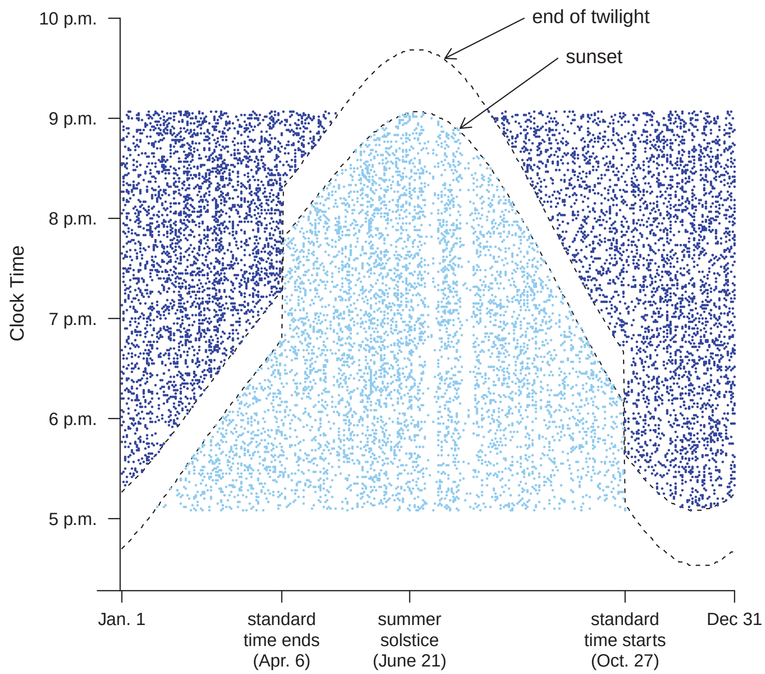
3.1 Intertwilight interval

The evening intertwilight interval is defined as clock times between the earliest and latest sunsets of the year; these are the hours for which it is light at some times of the year and dark at other times. Stops that took place before sunset are categorized as daylight stops. Stops that took place after the end of civil twilight are classified as darkness stops. Stops that took place during civil twilight are excluded in order to ensure a sharp distinction between daylight and darkness. The morning intertwilight period (only 4.5 percent of intertwilight stops) is defined analogously. Times for sunrise, sunset, beginning of civil twilight, and end of civil twilight are from the U.S. Naval Observatory.[14]

A scatterplot of the dates and times of stops is shown in figure 1. Note that the shift to and from daylight saving time causes discontinuities in the times that separate daylight stops from darkness stops and stretches the intertwilight period. The morning intertwilight period is two hours shorter than the evening period because the discontinuities instead compress the intertwilight interval.

[14]http://aa.usno.navy.mil/data/docs/RS_OneYear.php

Figure 1: Date and Time of Intertwilight Traffic Stops



Notes: The curved white band is the civil twilight interval. Stops below the white band occurred during daylight. Stops above the white band occurred during darkness. Stops during civil twilight are excluded from the analysis.

3.2 Discretionary searches

During the data collection, when a driver, vehicle, or passenger was searched, the officer was required to record which type of search took place and one of the following "authority to search" options: verbal or written consent, "observation of contraband," "officer safety," or "incident to arrest." The model applies only to searches the officer *decides* to undertake—discretionary searches. As Hernández-Murillo and Knowles (2004) point out, failing to make this distinction has important consequences. To some extent, the daylight-darkness comparison differences away the misclassification problem addressed by Hernández-Murillo and Knowles, but the Minneapolis data has the advantage of being able to separate discretionary and non-discretionary searches. However, because of the design of the data collection process, there is some ambiguity about whether certain searches were discretionary. I therefore use two definitions of discretionary searches, one that defines them narrowly, excluding the ambiguous categories, and one that takes a more expansive view of what searches might be discretionary.

I exclude searches in which only the passenger is searched because the race of passengers was not recorded. Consent searches and officer safety searches are clearly discretionary.

The "incident to arrest" and "contraband observed" categories are where the narrow and broad definitions differ. When an arrest takes place, the police officer is required to search the arrested individual, thus the search is considered "incident to arrest." There are complications, however. If the officer had probable cause to arrest the driver before conducting the search, the search can be legally considered "incident to arrest," even if an arrest does not take place. In State v. Bauman,[15] a Minnesota case that explicated this principle, the situation was described as follows:

> At the time [the officer] asked Bauman to step out of his car so [the officer] could search for identification, [the officer] had accumulated significant information. He knew that the driver had given the name and date of birth of an individual with a valid driver's license, but

[15] State v. Bauman no. C5-98-224, November 24, 1998.

> the registered owner of the vehicle had a suspended license. He also knew that the driver said he did not have his license with him, could not say how old he was, was unsure of his address, was unsure of the purported car owner's address, and could not tell the officer where he was going. On these facts, he had probable cause to arrest Bauman for providing false information to an officer.

This search was clearly discretionary, and such would be the case whenever an arrest did not take place and the search was undertaken under the legal authority of incident to arrest. Searches that fit this description are, therefore, considered discretionary under both definitions.

In the Bauman case the officer decided not to arrest the driver, but it is easy to imagine that the search could have turned up marijuana or stolen property. In that scenario, an arrest would have taken place, but the search itself would have been no less discretionary. Unfortunately, it is impossible to distinguish searches that fit this scenario from searches that took place *because* the driver was arrested—for example, was identified as the suspect in a prior crime. Thus the broader definition of discretionary search includes all searches classified as incident to arrest.

There is one further complication involving incident to arrest searches. When a vehicle is impounded, a search of the vehicle is mandatory, but impoundment was not an authority to search option on the form. Apparently, for impoundments, officers typically classified the search as "incident to arrest" (probably because this generally implies a mandatory search). Thus, I consider incident to arrest searches in which only a vehicle search took place to be non-discretionary under both definitions.[16]

[16]See CCJ/IRP (2003) for details. Except for five "contraband observed" searches, my narrow definition of discretionary searches exactly follows that report.

4 Empirical implementation

4.1 Stops

This section presents results from applying GR's method to 2002 data from Minneapolis. The analysis of stop rates draws heavily on Ritter and Bael (2009). As mentioned earlier, the test is based on an underlying difference-in-differences logic with darkness as the treatment. However, since the means involved (fraction of stops with white drivers and fraction with nonwhite drivers) must add to one, they are not independent, so GR's implementation uses logit regressions in which the dependent variable is an indicator of a nonwhite driver, and the key independent variable is an dummy for darkness.[17]

Table 2 reports results of regressions in which only stops from the morning and evening intertwilight periods are included. The first column reports estimates that use no additional controls. The second columns adds dummies for 16-minute intervals of clock time (16 minutes divides both intertwilight period almost exactly) to allow for possible intra-day changes. The third column also adds seasonal controls in the form of month dummies.

The results indicate that the share of stopped drivers who are black, Hispanic or nonwhite is between 4 and 7 percentage points lower during darkness, as predicted by Proposition 3. The magnitudes of the effects are substantial in light of the fact that the differentials must be generated from the relatively small share of stops in which the race of the driver can be observed ex ante. The magnitudes are also quite robust to the inclusion of time-of-day and month controls. There seems to be little doubt that there was racial bias in traffic stops during 2002.

[17] Although I follow GR's lead in using logits, all results are nearly identical using linear probability models.

Table 2: Stop rate tests

Sample	Average marginal effect of darkness		
Black and white drivers	-0.052^{***}	-0.068^{***}	-0.058^{**}
(black = 1, N = 10,567)	(0.010)	(0.011)	(0.021)
Hispanic and white drivers	-0.042^{***}	-0.048^{***}	-0.039^{*}
(Hispanic = 1, N = 7,145)	(0.010)	(0.011)	(0.020)
Nonwhite and white drivers	-0.049^{***}	-0.061^{***}	-0.055^{***}
(nonwhite = 1, N = 12,908)	(0.009)	(0.010)	(0.019)
Time-of-day controls	No	Yes	Yes
Month controls	No	No	Yes

Significance: *** = 0.01, ** = 0.05, * = 0.1. Average marginal effects from logit regressions using stops from morning and evening intertwilight periods. Bootstrap standard errors based on 500 replications. Time-of-day controls are dummies for 16 minute intervals.

4.2 Searches

Proposition 4 provides additional testable implications of the model: darkness should raise the probability of discretionary searches of nonwhite drivers and lower the probability for white drivers. Table 3 reports results of regressing an indicator for a discretionary search on an indicator for darkness and controls. Stops leading to non-discretionary searches are excluded. The impacts of darkness are nearly all insignificant and many signs are opposite those predicted. The coefficients for white drivers in columns 1 and 2 are statistically significant and their signs are consistent with predictions. However, these effects are not robust to inclusion of seasonal controls or to the broader definition of discretionary.

As mentioned above, GR's test potentially has low power because only cases in which the police officer observed the driver's race before the stop help to identify the effect of darkness. That is not an issue for these tests, which would have low power in the opposite situation—if officers could usually identify the driver's race ex ante. The search rate tests could have low power if the J thresholds for different race drivers were close together, so that darkness would not change selection into the population of stopped drivers very much. The results on stop rates indicate that is not the case, however.

Table 3: Impact of darkness on discretionary search rate

Controls	Narrow definition			Broad definition		
Black drivers	−0.012	−0.015	−0.014	−0.019	−0.023	−0.035
	(0.012)	(0.013)	(0.023)	(0.012)	(0.014)	(0.026)
		N = 4,487			N = 4,975	
Hispanic drivers	0.019	0.018	0.044	0.042*	0.047*	0.045
	(0.020)	(0.021)	(0.040)	(0.023)	(0.026)	(0.046)
		N = 1,139			N = 1,583	
Nonwhite drivers	−0.006	−0.007	−0.001	−0.007	−0.007	−0.009
	(0.009)	(0.010)	(0.019)	(0.010)	(0.011)	(0.021)
		N = 6,455			N = 7,225	
White drivers	−0.011*	−0.014**	−0.001	−0.007	−0.012	0.000
	(0.006)	(0.006)	(0.013)	(0.008)	(0.008)	(0.014)
		N = 5,260			N = 5,427	
Time-of-day controls	No	Yes	Yes	No	Yes	Yes
Month controls	No	No	Yes	No	No	Yes

Significance: *** = 0.01, ** = 0.05, * = 0.1. Dependent variable is a dummy for a discretionary search; sample excludes stops resulting in non-discretionary searches. Average marginal effects from logit regressions using stops from morning and evening intertwilight periods. Bootstrap standard errors based on 500 replications. Time-of-day controls are dummies for 16 minute intervals.

Table 4: Stop rate tests, driving violations only

Sample	Average marginal effect of darkness		
Black and white drivers	-0.056^{***}	-0.072^{***}	-0.077^{***}
(black = 1, N = 6,334)	(0.013)	(0.014)	(0.024)
Hispanic and white drivers	-0.033^{***}	-0.044^{***}	-0.041^{*}
(Hispanic = 1, N = 4,426)	(0.012)	(0.014)	(0.024)
Nonwhite and white drivers	-0.052^{***}	-0.067^{***}	-0.074^{***}
(nonwhite = 1, N = 7,592)	(0.011)	(0.012)	(0.023)
Time-of-day controls	No	Yes	Yes
Month controls	No	No	Yes

Significance: *** = 0.01, ** = 0.05, * = 0.1. Average marginal effects from logit regressions using stops from morning and evening intertwilight periods. Bootstrap standard errors based on 500 replications. Time-of-day controls are dummies for 16 minute intervals.

4.3 Sensitivity analysis

This section considers two variations on the tests reported in the previous sections. First, some types of equipment violations might be less visible during darkness, while others, such as burned-out headlights, would be more visible. Equipment violations are likely correlated with income and, therefore race. The same might be true of expired license tags. The proportion of stopped drivers who are nonwhite might drop after dark for these reasons, rather than racial bias. I therefore repeated the estimation with a sample including only stops triggered by driving violations. The results from using this restricted sample, reported in tables 4 and 5, are very similar to those from the full sample. In particular, there is strong evidence of racial disparity in stops and no evidence in searches.

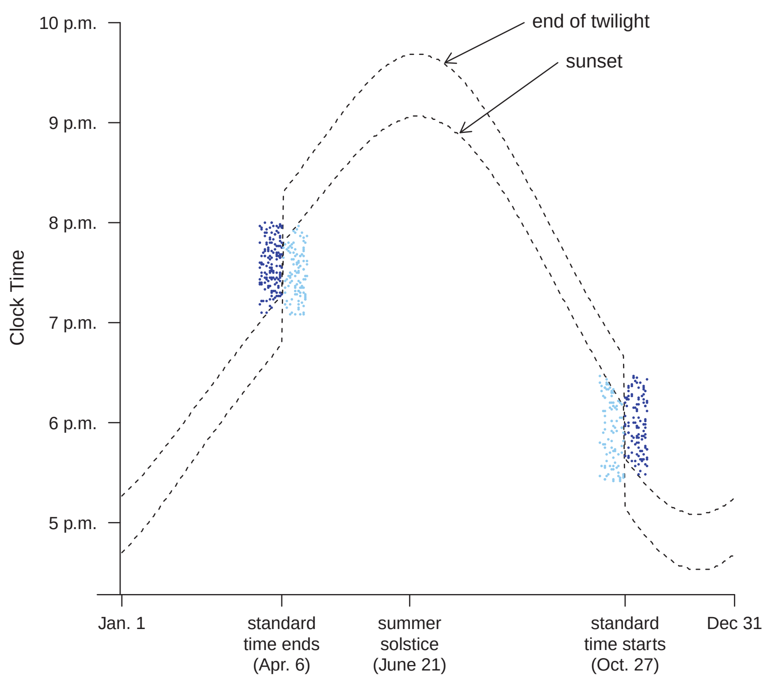
Table 5: Impact of darkness on discretionary search rate, driving violations only

Controls	Narrow definition			Broad definition		
Black drivers	−0.004	−0.002	0.016	−0.019	−0.023	−0.035
	(0.015)	(0.016)	(0.031)	(0.012)	(0.014)	(0.026)
		$N = 2{,}449$			$N = 2{,}703$	
Hispanic drivers	0.026	0.029	0.048	0.042	0.023	0.066
	(0.026)	(0.030)	(0.057)	(0.023)	(0.026)	(0.070)
		$N = 693$			$N = 812$	
Nonwhite drivers	0.005	0.004	0.026	−0.007	−0.007	−0.009
	(0.011)	(0.014)	(0.024)	(0.010)	(0.011)	(0.021)
		$N = 3{,}518$			$N = 3{,}915$	
White drivers	0.000	−0.002	0.000	−0.007	−0.012	0.000
	(0.006)	(0.006)	(0.014)	(0.008)	(0.008)	(0.014)
		$N = 3{,}464$			$N = 3{,}548$	
Time-of-day controls	No	Yes	Yes	No	Yes	Yes
Month controls	No	No	Yes	No	No	Yes

Significance: *** = 0.01, ** = 0.05, * = 0.1. Dependent variable is a dummy for a discretionary search; sample excludes stops resulting in non-discretionary searches. Average marginal effects from logit regressions using stops from morning and evening intertwilight periods. Bootstrap standard errors based on 500 replications. Time-of-day controls are dummies for 16 minute intervals.

Second, although the estimates in sections 4.1 and 4.2 include seasonality controls, an alternative approach is to take advantage of the switches to and from daylight saving time—a natural experiment within a natural experiment. Near these switches there is a about an hour of clock time that switches between darkness and daylight from one day to the next, and a wider range of clock time that switches over the course of a few days. To take advantage of this situation I narrowed the sample to stops that occurred within two weeks of one of the time changes, a total of only 470 stops. The resulting evening sample is illustrated in figure 2. The results reported in tables 6 and 7 are quite similar to those reported above. There is clear evidence of racial bias in stops, but no evidence of bias in search decisions. Some of the marginal effects in the search rate tests are statistically significant, but these all have signs opposite those predicted by Proposition 4.

Figure 2: Date and Time of Traffic P.M. Stops in DST-Switches Sample



Notes: The curved white band is the civil twilight interval. Stops below the white band occurred during daylight. Stops above the white band occurred during darkness. Stops during civil twilight are excluded from the analysis.

Table 6: Stop rate tests using only stops near time changes

Sample	Marginal effect of darkness	
Black and white drivers	-0.171^{***}	-0.171^{***}
(black = 1, $N = 376$)	(0.053)	(0.02)
Hispanic and white drivers	-0.037	-0.030
(Hispanic = 1, $N = 276$)	(0.053)	(0.054)
Nonwhite and white drivers	-0.120^{***}	-0.115^{**}
(nonwhite = 1, $N = 470$)	(0.046)	(0.050)
Time-of-day controls	No	Yes

Significance: *** = 0.01, ** = 0.05, * = 0.1. Average marginal effects from logit regressions using stops from morning and evening intertwilight periods. Bootstrap standard errors based on 500 replications. Time-of-day controls are dummies for 16 minute intervals. Time-of-day dummies capture any difference between spring and fall because spring and fall times do not overlap (see Figure 2).

Table 7: Marginal effect of darkness on discretionary search rate using only stops near time changes

Controls	Narrow definition		Broad definition	
Black drivers	-0.038	-0.037	-0.116^{*}	-0.148^{**}
	(0.057)	(0.069)	(0.069)	(0.075)
	$N = 143$		$N = 158$	
Hispanic drivers	-0.051	0.054	0.021	-0.020
	(0.078)	(0.108)	(0.103)	(0.121)
	$N = 54$		$N = 61$	
Nonwhite drivers	-0.047	-0.058	-0.076	-0.097^{*}
	(0.043)	(0.045)	(0.048)	(0.054)
	$N = 224$		$N = 247$	
White drivers	0.011	0.041	0.066^{**}	0.087^{**}
	(0.020)	(0.022)	(0.030)	(0.032)
	$N = 196$		$N = 206$	
Time-of-day controls	No	Yes	No	Yes

Significance: *** = 0.01, ** = 0.05, * = 0.1. Dependent variable is a dummy for a discretionary search; sample excludes stops resulting in non-discretionary searches. Average marginal effects from logit regressions using stops from morning and evening intertwilight periods. Bootstrap standard errors based on 500 replications. Time-of-day controls are dummies for 16 minute intervals. Time-of-day dummies capture any difference between spring and fall because spring and fall times do not overlap (see Figure 2).

5 Discussion

The race or ethnicity of a driver may be informative about the probability he has committed a crime. That fact will generate a disparity in traffic stops, if police statistically discriminate. However, the race or ethnicity of a driver is often not known until after a traffic stop takes place. A police officer can then use race or ethnicity to update his or her assessment of the driver. On average, during darkness a stop reveals more information because the driver's skin color is less likely to be observed before the stop. Similarly, an officer can act on a taste for discrimination when race is revealed during a stop, if it was not known before. It is this simple updating that motivates the search rate tests introduced in this paper.

Before discussing the Minneapolis results, it is worth touching briefly on different outcomes that could arise in different data. In particular, it is possible that the data would be consistent with all of the predictions of Propositions 3 and 4, suggesting the presence of either statistical discrimination or taste for discrimination, but not distinguishing between the two. To make that distinction it would be necessary to employ a "guilt" outcome such as the hit rate for searches. Subject to data availability and the caveats discussed in section 1.2, KPT's, AF's, or Antonovics and Knight's methods could supplement the tests used above.

One important contribution of this paper, however, is to highlight that applying those tests in the wrong context can be misleading. For example, hit rates are very unequal in the Minneapolis data (CCJ/IRP, 2003), but that cannot imply prejudice if police officers do not use information about race in the search decision (as the results for Minneapolis indicate).

5.1 An implausible interpretation of the results

The stop-rate and search-rate tests clearly indicate that, while there was racial bias in Minneapolis traffic stops during 2002, the nature of that bias is not consistent with statistical discrimination. Neither is the pattern of the results consistent with police officers who act optimally on a taste for discrimination.

One possible interpretation of this pattern is that police officers are simply bad Bayesians: They recognize that skin color conveys useful law enforcement information and this is reflected in stop rates, but correctly updating their beliefs after a stop takes place imposes an excessive cognitive burden. There is, after all, plenty of evidence that human beings have very bad intuitions about Bayes' rule.

This bad Bayesians interpretation is not plausible, however. In practical terms

Proposition 4 is just an aggregate implication of individual police officers using information about skin color revealed by the stop in exactly the same way they are hypothesized (under this interpretation) to use skin color information available before the stop. There is no reason to think the cognitive burden is higher in the search decision than in the stop decision. Bad Bayesians might not use information optimally, but the results suggest that officers do not respond to it at all at the search stage, where it is *always* available. Failing to update at all when informative data is literally in front of one's nose cannot be construed as even crude Bayesian reasoning.

5.2 A plausible interpretation of the results

On the other hand, the behaviors and attitudes labeled by society as prejudice need not be consistent with the kind of optimization implied by taste-based discrimination.

The decision to stop a vehicle is typically made under time pressure and ambiguity (unless it is motivated by a major traffic violation). Subsequent steps in a traffic stop are taken much more deliberately. Research on decision processes reveals sharply different outcomes between decisions made rapidly on the basis of intuition and those to which conscious thought has been applied (Kahneman, 2011). In terms of the metaphor employed by Kahneman, the stop decision must frequently be made by System 1, while System 2 can play a role in the decision of whether to initiate a search.

"Implicit discrimination," described by Bertrand, Chugh, and Mullainathan (2005), can be understood in this way. Implicit discrimination is hypothesized to be a consequence of unconscious mental associations of race with actions or characteristics and thus would be a characteristic of System 1. Bertrand et al. point out that decisions made under time pressure and involving a high level of ambiguity encourage the exercise of implicit discrimination. Certainly the decision to stop a vehicle is often made under time pressure and ambiguity. One would thus expect implicit discrimination to lead to evidence of bias in stop decisions.

Subsequent steps in a traffic stop are taken much more deliberately, however. There is certainly residual ambiguity in a decision to search, but by the time a search decision is made, the officer will typically have gathered much more information about the driver and vehicle: both are seen at close range, license and vehicle registration are checked, questions can be asked. The officer thus has several minutes to consider the search decision, and System 2 can be engaged. These conditions of deliberation would tend to mitigate the expression of implicit discrimination, which is consistent with failing discover evidence that race is used in search decisions.[18]

Although conclusions about implicit discrimination must be tentative, the novel methods introduced in this paper do reach an important conclusion: both of the "usual suspects"—statistical discrimination and Becker-type taste-based discrimination—are clearly inconsistent with the data. Researchers must look deeper to understand how race is used by police in this context.

[18] Any conclusion about implicit discrimination based on the present research must be tentative because other hypotheses also predict different results for searches than for stops. For example, if race were used only to help judge the probability a driver is a known suspect, there might be racial bias in stops, but since the uncertainty would be fully resolved by a license check, there would be no reason for any bias in discretionary searches. Or officers could simply have a taste for discrimination in stops, but not in searches (there being no reason to expect prejudice to be internally consistent).

References

Antonovics, Kate and Brian C. Knight, "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics,* Vol. 91, No. 1, (2009).

Anwar, Shamena and Hanming Fang, "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Revew,* Vol. 96, No. 1 (2006).

Barbe, Philippe and William C. Horrace, "A Critical Reanalysis of Maryland State Police Searches," *The American Statistician,* Vol. 66, No. 1 (2012).

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan, "Implicit Discrimination," *American Economic Review,* Vol. 95, No. 2 (2005).

Close, Billy R. and Patrick L. Mason, "After the Traffic Stops: Officer Characteristics and Enforcement Actions," *The B.E. Journal of Economic Analysis & Policy,* Vol. 6, No. 1 (2006).

Council on Crime and Justice and Institute on Race and Poverty, "Minnesota Statewide Racial Profiling Report: Minneapolis Police Department," Report to the Minnesota Legislature, September 22, 2003.

Grogger, Jeffrey, and Greg Ridgeway, "Testing for Racial Profiling in Traffic Stops from behind a Veil of Darkness," *Journal of the American Statistical Association,* vol. 101, no. 475, (September 2006).

Hernández-Murillo, Rubén and John Knowles, "Racial Profiling or Racist Policing? Bounds Tests in Aggregate Data," *International Economic Review,* Vol. 45, No. 3 (2004).

Kahneman, Daniel, *Thinking, Fast and Slow,* Farrar, Straus, and Giroux, 2011.

Knowles, John, Nicola Persico, and Petra Todd, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy,* Vol. 109, No. 2 (2001).

Myers, Samuel L., Jr., "Analysis of Racial Profiling as Policy Analysis," *Journal of Policy Analysis and Management,* Vol. 21, No. 2, 287-300 (2002).

Ritter, Joseph A. and David Bael, "Detecting Racial Profiling in Minneapolis Traffic Stops: A New Approach," *CURA Reporter,* Spring/Summer 2009.

Schell, Terry, Greg Ridgeway, Travis L. Dixon, Susan Turner, and K. Jack Riley, "Police-Community Relations in Cincinnati: Year Three Evaluation Report," Rand Corporation Technical Report 535, 2007.

Appendix: Proof of Proposition 4

I suppress both the time-of-day notation and the traffic-enforcement benefit, A. When stopping an n driver, the police officer might or might not have known he was an n driver. Thus, for any $t \in T$ and A,

$$\begin{aligned}
&\Pr(\text{search}|\text{stop}, n, \ell = 1) - \Pr(\text{search}|\text{stop}, n, \ell = 0) \\
&\qquad = \delta \Pr(\text{search}|\text{stop}, n, n) + (1-\delta)\Pr(\text{search}|\text{stop}, u) \\
&\qquad\quad - [\eta \Pr(\text{search}|\text{stop}, n, n) + (1-\eta)\Pr(\text{search}|\text{stop}, u)] \\
&\qquad = (\delta - \eta)\,[\Pr(J^* > J_n^* | J > J_n) - \Pr(J^* > J_n^* | J > J_u)] \\
&\qquad = (\delta - \eta)\left[\int_{J_n}^{\infty} \frac{\Pr(J^* > J_n^*|J)}{1 - F(J_n)}\,dF(J) - \int_{J_u}^{\infty} \frac{\Pr(J^* > J_n^*|J)}{1 - F(J_u)}\,dF(J)\right] \\
&\qquad = (\delta - \eta)\left[\frac{1}{1 - F(J_n)} \int_{J_n}^{J_u} \Pr(J^* > J_n^*|J)\,dF(J)\right. \\
&\qquad\quad \left. + \left(\frac{1}{1 - F(J_n)} - \frac{1}{1 - F(J_u)}\right) \int_{J_u}^{\infty} \Pr(J^* > J_n^*|J)\,dF(J)\right].
\end{aligned}$$

The first integrand is pointwise less than or equal to $\Pr(J^* > J_n^*|J_u)$ given the stochastic dominance assumption (Assumption 6). Similarly, the second integrand is pointwise greater than or equal to $\Pr(J^* > J_n^*|J_u)$, while the quantity in parentheses is negative. Therefore,

$$\begin{aligned}
&\Pr(\text{search}|\text{stop}, n, \ell = 1) - \Pr(\text{search}|\text{stop}, n, \ell = 0) \\
&\qquad \leq (\delta - \eta)\Pr(J^* > J_n^*|J_u)\left[\int_{J_n}^{J_u} \frac{1}{1 - F(J_n)} dF(J)\right. \\
&\qquad\quad \left. + \left(\frac{1}{1 - F(J_n)} - \frac{1}{1 - F(J_u)}\right) \int_{J_u}^{\infty} dF(J)\right] \\
&\qquad = (\delta - \eta)\Pr(J^* > J_n^*|J_u)\left[\frac{F(J_u) - F(J_n)}{1 - F(J_n)} - \frac{F(J_u - F(J_n))}{1 - F(J_n)}\right] \\
&\qquad = 0.
\end{aligned}$$

The previous inequality holds for every $t \in T$. Averaging over the intertwilight interval, T yields

$$\int_{t\in T} \left[\Pr(\text{search}|\text{stop}, n, \ell = 1, t) - \Pr(\text{search}|\text{stop}, n, \ell = 0, t)\right] g(t, n)\, dt < 0, \tag{A-1}$$

where $g(t, n) = N(t)/\int_{s\in T} N(s)\, ds$ is the density of observations of n drivers by police officers at t. Since inequality (A-1) holds for any A, integrating over the distribution of A produces the inequality in part (i) of the proposition. For w drivers,

$$\begin{aligned}
\Pr(\text{search}|\text{stop}, &w, \ell = 1) - \Pr(\text{search}|\text{stop}, w, \ell = 0)\\
&= \delta \Pr(\text{search}|\text{stop}, w, w) + (1 - \delta) \Pr(\text{search}|\text{stop}, u)\\
&\quad - \left[\eta \Pr(\text{search}|\text{stop}, w, w) + (1 - \eta) \Pr(\text{search}|\text{stop}, u)\right]\\
&= (\delta - \eta) \left[\Pr(J^* > J^*_w | J > J_w) - \Pr(J^* > J^*_w | J > J_u)\right]\\
&= (\delta - \eta) \left[\int_{J_w}^{\infty} \frac{\Pr(J^* > J^*_w|J)}{1 - F(J_w)}\, dF(J) - \int_{J_u}^{\infty} \frac{\Pr(J^* > J^*_w|J)}{1 - F(J_u)}\, dF(J)\right]\\
&= (\delta - \eta)\Bigg[\frac{1}{1 - F(J_u)} \int_{J_u}^{J_w} \Pr(J^* > J^*_w|J)\, dF(J)\\
&\quad + \left(\frac{1}{1 - F(J_w)} - \frac{1}{1 - F(J_u)}\right) \int_{J_w}^{\infty} \Pr(J^* > J^*_w|J)\, dF(J)\Bigg].
\end{aligned}$$

The numerators of the both integrands are pointwise greater than or equal to $\Pr(J^* > J^*_w|J_u)$. In this case the quantity in parentheses is positive. Therefore

$$\begin{aligned}
\Pr(\text{search}|\text{stop}, &w, \ell = 1) - \Pr(\text{search}|\text{stop}, w, \ell = 0)\\
&\geq (\delta - \eta) \Pr(J^* > J^*_w|J_u) \left[\frac{2\big(F(J_w) - F(J_u)\big)}{1 - F(J_u)}\right]\\
&> 0,
\end{aligned}$$

Again, the inequality holds for every $t \in T$, and averaging over T and A proves part (ii).