Out-of-Place and In-Place Policing: An Examination of Traffic Stops in Racially Segregated
Philadelphia

Lance Hannon
Malik Neal
Alex R. Gustafson

(Forthcoming in *Crime and Delinquency*, the most recent version of this article is published here:
<https://doi.org/10.1177/0011128720926122>.)

Out-of-Place and In-Place Policing: An Examination of Traffic Stops in Racially Segregated Philadelphia

Abstract

It is commonly argued that Black people may be more likely to be stopped by the police in majority White neighborhoods due to a natural tendency to first observe and then scrutinize that which seems out of the ordinary. Anecdotal evidence of police officers appearing equally drawn to White people in predominantly Black neighborhoods is sometimes presented to suggest that the phenomenon is race neutral. Motivated by such narratives, we examine the extent to which Black versus White racial categorization encourages police scrutiny in out-of-place and in-place contexts. Applying the veil-of-darkness and vehicle search threshold tests, we find that in-place or out-of-place, being seen as White is always an advantage in Philadelphia.

Introduction

It is unlawful for police officers to initiate investigatory stops of vehicles based solely on the driver's race or ethnicity (Kennedy, 1997). However, as long as it is only one component of reasonable suspicion, the courts have traditionally allowed police officers to consider whether a motorist is "out of place" given the empirical reality of persistently high levels of residential segregation in the United States (Kennedy, 1997; Russell Brown, 2008). Underlying such authorization is the recognition that good police work frequently requires officers to be observant of outliers.

A clear example of how the courts have conceptualized the issue of race-out-of-place appears in *State v. Dean* (Arizona, 1975). In this case, Dean alleged that the detention of his vehicle was invalid because, as admitted by the investigating officer, part of the reason why Dean was initially seen as suspicious was that he was a Latinx male in a predominantly White neighborhood. The Arizona Supreme Court ruled that the stop was lawful because the perceived mismatch between Dean's ethnicity and the racial composition of the area was only part of the totality of circumstances for reasonable suspicion. The Court concluded:

That a person is observed in a neighborhood not frequented by persons of his ethnic background is quite often a basis for an officer's initial suspicion. To attempt by judicial fiat to say he may not do this ignores the practical aspects of good law enforcement. While detention and investigation based on ethnic background alone would be arbitrary and capricious and therefore impermissible, the fact that a person is obviously out of place in a particular neighborhood is one of several factors that may be considered by an officer and the court in determining whether an investigation and detention is reasonable and therefore lawful.

This common line of argument is frequently described as the “out-of-place doctrine” (Russell Brown, 2008; Thompson, 1999).

When explaining the difference between good police work and bias-based enforcement, police officers also commonly reference the practical need to react to that which is out of the ordinary. As detailed in Glover’s (2007) ethnographic research, officers regularly state that race-out-of-place individuals “stand out more than anything” (p. 243) and this is true for White people observed in predominantly Black neighborhoods, not just racial minorities in White spaces. Glover argues that police narratives depicting out-of-place White suspects tacitly promote the notion that policing is race neutral (in the sense that White people and Black people are portrayed as equally subject to distrust in certain contexts). While the officers in Glover’s sample acknowledged that an individual’s race interacts with the racial composition of the neighborhood to either heighten or lessen suspicion, the officers rarely discussed the case of minorities in White neighborhoods without first mentioning the example of the “White boy in a no White boy zone” (Glover, 2007, p. 242).

Supporting Glover’s (2007) conclusions, a cursory review of anonymous law enforcement boards reveals similar discourse regarding valid uses of race for determining reasonable suspicion. For example, consider this verbatim entry on forum.officer.com concerning the difference between racial profiling and criminal profiling:

Racial profiling, IMHO, is using someones race or ethnic catagory solely to determine if that person would commit a particular type of crime. Many people state that racial profile is the bases for X number of (insert any one particular race) is being stopped more frequently than (insert a different race). What people fail to see (again, IMHO) is you have to consider the demographics being questioned in those scenarios. For example; I

work a high crime, high drug area of town (commonly referred to as the projects). It is a matter of fact, not opinion, that the majority of persons in this area are black. When I'm patrolling in the area and I see a white person, I am going to stop and speak to them. This isn't racial profiling. 1) this is a predominantly black community. 2) this is a high drug area. 3) my training and experience is that white men come to X area to purchase X drug. 4) I know he doesn't belong in the area, meaning he doesn't live there. This is my zone...you learn to know who does and doesn't live there. It's a matter of good policing. The man being white isn't the ONLY reason I stopped him. Did he get my attention because he's white...absolutely he did.

In this example, in line with legal precedent, the anonymous author emphasizes that while race-out-of-place is often the initial factor drawing an officer's attention, it is only one factor in the totality of circumstances defining reasonable suspicion and it applies to White people too.

Withrow's (2004) theory of contextual attentiveness formalizes this common perspective among law enforcement personnel as a potential explanation for racial disparities in policing. Withrow notes that the police are especially focused on deviations from the norm and can always find a probable-cause reason for stopping a vehicle when they see a driver that is noticeably incongruent with local context. Arguing that both of the following situations are "equally inconsistent" (Withrow, 2004, p. 359), Withrow poses the question: "Is a White man driving in a neighborhood populated predominately by Black residents as likely to be stopped and searched as a Black man driving in a neighborhood populated predominately by White residents?" (2004, p. 345). If so, Withrow concludes that rather than "driving while Black," the real issue could be "driving while different" (Withrow, 2004, p. 361).

The present study aims to evaluate whether there is racial parity in the degree to which being out-of-place versus in-place matters for the relative likelihood of being stopped and searched by the police. Focusing on Philadelphia, a city that has been described as hyper-segregated (Massey & Denton, 1998), we use publicly available data from the American Community Survey and the Philadelphia Police Department. Our analyses employ a third-generation benchmarking technique known as the veil-of-darkness test to help ensure that any observed racial disparities in stop rates are due to police officer discretionary judgments rather than differences in traffic violation rates or vehicle characteristics (Pierson et al., 2019; Smith et al., 2019). Additionally, we utilize stop outcome tests that do not require an external benchmark to assess the extent to which fitting in demographically matters for avoiding unproductive vehicle searches.

Previous Studies

Because there are many excellent reviews of the vast social science literature on biased-based policing (Harris, 2002; Smith et al., 2019; White & Fradella, 2016), we focus our attention on studies that (1) specifically aim to assess whether out-of-place policing occurs equally for White and Black people and (2) avoid reliance on the debunked premise that the population traveling through an area accurately resembles the demographic composition of local residents (Hannon, 2019; Tillyer, Engel, & Wooldredge, 2008).

Overall, existing research on this topic offers some support for the notion that out-of-place policing runs in both directions, although this research does not indicate that it runs equally in both directions. In particular, there is some evidence that in certain cities Black search rates are higher in predominantly White areas, and White search rates are higher in predominantly

Black areas (Novak & Chamlin, 2008; Rojek, Rosenfeld, & Decker, 2012). We note that this scholarship does not claim to be conclusive because there are also studies focused on other areas reporting that the size of the local Black population is positively associated with higher search rates for both Black and White drivers (Close & Mason, 2007) or not meaningfully associated with search rates for either set of drivers (Carroll & Gonzalez, 2014). Furthermore, we note that when drawing conclusions about out-of-place policing, it is very important to keep an eye on which groups are being compared in the analytic techniques employed in this body of research. More specifically, the interpretation of results can vary dramatically depending on the implicit contrast group in the models (e.g., is the reported result comparing how White individuals are treated relative to Black individuals in various contexts or is the result simply comparing differently situated White individuals?). We describe some of the nuanced complexity in this literature below.

Meehan and Ponder (2002) offer one of the earliest sophisticated tests of out-of-place policing. One of the unique features of this study is that it utilized data regarding computer queries initiated by suburban police officers surveying suspicious vehicles while out on patrol. Examining this data, Meehan and Ponder (2002) found that as Black drivers moved toward the overwhelmingly White center of the suburban jurisdiction (and away from a bordering majority Black city), they faced a significant increase in the probability of being subject to an electronic query. Interestingly, Meehan and Ponder (2002, p. 417) noted that for White drivers, the chance of being queried was roughly equal throughout the jurisdiction (regardless of proximity to the bordering majority Black city).

Novak and Chamlin (2012) examined traffic stop data for Kansas City, Missouri and provided a novel conceptual framework that combined insights from the racial threat perspective

with empirical work on out-of-place policing. They noted (2012, p. 275) that their “most intriguing finding” involved “the conditional effect of the racial composition of the beat on search rates.” In particular, they found that the percentage Black in a police beat was significantly positively associated with the search rate for White drivers, while the coefficient was smaller and statistically insignificant (but still positive) for the regression model predicting the search rate for Black drivers. Therefore, Novak and Chamlin’s (2012) analyses pointed to an out-of-place policing effect that was only statistically discernible for White drivers. Put differently, while White drivers experienced a clear benefit from being in-place, Black drivers did not. They instead experienced a constantly elevated probability of being searched relative to White motorists that was irrespective of in-place or out-of-place status. Indeed, the coefficients in Novak and Chamlin’s (2012) regression models indicated that while the predicted White search rate got closer to the Black search rate as the percentage Black increased, the Black search rate was still higher than the White search rate in beats that were 90% Black (holding constant other factors at zero).

In a study that tested a variety of important theoretic propositions simultaneously, Rojek, Rosenfeld, and Decker (2012) utilized traffic stop data for the city of St. Louis. While the city has roughly equal numbers of White and Black residents, Rojek, Rosenfeld, and Decker (2012) noted that St. Louis is highly segregated, with some police districts being populated almost entirely with Black residents. For the city as a whole, Rojek, Rosenfeld, and Decker (2012) reported that the Black search rate was about 70% higher than the White search rate (6.3% versus 3.7%). However, their logistic regression results indicated considerable variation in the relative likelihood of a search that was contingent on the particular combination of the race of the officer and the race of the driver as well as the racial composition of the police district where the

stop took place. Ultimately, Rojek, Rosenfeld, and Decker (2012) concluded that in districts where Black people were not in the majority, stops involving White officers and Black drivers were the ones that were most likely to end in a search. Furthermore, in majority Black districts, searches appeared to be most likely for stops of White drivers by White officers. Thus, Rojek, Rosenfeld, and Decker's research suggests that out-of-place policing occurs for both White and Black motorists, although the analyses were not aimed toward quantifying the degree to which this happens evenly to both sets of drivers and the particular structure and complexity of the models make after-the-fact calculations unhelpful (Black officer-White driver is the omitted contrast group in the equations).

Carroll and Gonzalez (2014) expanded this line of inquiry beyond searches to also include frisks. They reasoned that because frisks require a lower standard of justification (reasonable suspicion) than searches (probable cause), racial bias will be more evident for vehicle stops with frisks than searches. Their analyses of traffic stop data from the Rhode Island State Police supported this hypothesis. More important, Carroll and Gonzalez (2014) examined the conditioning impact of the racial composition of the location where the stop took place and found that Black drivers were significantly more likely to be frisked when they were stopped in the predominately White townships outside of the Providence area. Interestingly, similar to Meehan and Ponder's (2002) findings, Carroll and Gonzalez (2014) reported that the White frisk rate did not vary much with local racial context and was never higher than the Black frisk rate. That is, (1) out-of-place policing appeared to be much more of a phenomenon for Black drivers than White drivers and (2) even though the estimated race-specific frisk rates were closer outside of White suburbia, being perceived as Black was never an advantage relative to being perceived as White. Carroll and Gonzalez (2014) suggested that this asymmetry may be due to different

stereotypes associated with out-of-place White people versus out-of-place Black people. The former may be perceived as a hapless shopper of illicit goods while the latter may be seen as outright dangerous.

Like Carroll and Gonzalez (2014), a significant component of Levchak's (2017) comprehensive study focused on frisks as a highly discretionary stop outcome. Examining pedestrian stop data for New York City, Levchak (2017) estimated the degree to which the racial composition of a police district influenced the association between individual race and the probability of a frisk occurring during a stop. In this case, not only was the Black frisk rate always estimated to be higher than the White frisk rate, the disparity between the two actually grew as the percentage Black in the police district increased. Levchak concluded (2017, p. 397): "Thus, compared to whites, blacks are more likely to be frisked in precincts with large concentrations of African Americans—suggesting that whites may not experience out-of-place policing here."

Data and Methods

Associated with the settlement of *Bailey v. City of Philadelphia* (2011), a case that focused on racial disparities in pedestrian stop-and-frisk practices, the Philadelphia Police Department currently publishes comprehensive data on investigative detentions occurring within city limits.¹ Different from other jurisdictions (e.g., Chicago), all routine traffic stops occurring on local streets were deemed investigative detentions and thus included alongside pedestrian stops in this large public dataset ($N > 2$ million stops). The vast majority of vehicle stops (approximately 95%) were formally justified based on motor vehicle code violations (as opposed to vehicle matching flash description or other rationales) and arrests for any offense were rare

(just 2% for Black motorists and 2% for White motorists). Highway vehicle stops in Philadelphia were not included in the data because those are primarily handled by the Pennsylvania State Police.

Pierson et al. (2019) recently conducted a number of illustrative analyses focused on the vehicle portion of Philadelphia's public stop dataset (<https://openpolicing.stanford.edu/tutorials/>). Their analyses demonstrated considerable citywide racial disparities in stop rates, search rates, and contraband recovery rates from searches. In particular, their examination of the data for 2017 suggested that Black people in Philadelphia were stopped and searched at higher than expected rates, especially relative to White people. While most of Pierson et al.'s (2019) analyses were rudimentary (as their stated intention was to provide an introduction to analyzing the data), they also included two more advanced techniques for uncovering potential racial discrimination: (1) the veil-of-darkness test and (2) the search threshold test. The results of these more rigorous tests also indicated significant racial bias against Black drivers relative to White drivers for Philadelphia as a whole.

We build on Pierson et al.'s (2019) results by analyzing whether the overall findings associated with these tests are different when the data are disaggregated to a more local level.² Considering the longstanding hyper-segregation of Philadelphia's White and Black communities (Massey & Denton, 1998) and the importance of evaluating the symmetry/asymmetry of out-of-place policing, we focus our analytic attention on assessing racial disparities in predominantly White vs. predominantly Black areas. To do this, we match residential demographic estimates from the 2012-2016 American Community Survey (U.S. Census Bureau, 2019) to police district boundaries (police district indicators are included in the public stop data). More specifically, the longitude and latitude coordinates for the population center of each census block group were

merged with spatial parameters for Philadelphia's 21 police districts (there were 1336 census block groups in city limits). Ultimately, this procedure indicated that there were 9 districts that were majority Black and 7 districts that were majority White. Of the 5 remaining districts, 2 were largely Latinx and the other 3 were racially and ethnically diverse (i.e., not simply an equal mix of Black and White residents).³

We focus our analyses on the years 2015 through 2018. At the time of this writing, the data for 2019 were not complete and available information suggests that the year was characterized by a unique rise in vehicle stops (potentially reflecting a recent shift in policy, see Melamed, 2019). The data for 2014, the first year the data monitoring system was implemented, were also incomplete and seemed to indicate an escalation in the recording of stops (vehicle and pedestrian investigations rose very closely together in the first several months). Figure 1 depicts the available monthly data with a best-fitting cubic regression line through the vehicle stop observations.

The number of vehicles stops occurring between 2015 and 2018 was over 1 million, with the vast majority taking place in areas that were predominantly Black. American Community Survey estimates suggest that the City of Philadelphia's non-Latinx Black and White populations were each approximately 40% of the total residential population. The predominantly Black police districts used in our main analyses were, on average, 74% non-Latinx Black and 15% non-Latinx White. Conversely, the predominantly White police districts in our main analyses were, on average, 70% non-Latinx White and 12% non-Latinx Black.⁴

While there are many studies utilizing the veil-of-darkness test, ours is the first to employ it as a means for examining whether being classified as the same race as most local residents is equally advantageous for White and Black motorists at the decision-to-stop stage. Likewise, we

join only a handful of existing studies (e.g., Carroll & Gonzalez, 2014) examining the outcome of vehicle searches for contraband as a way to ascertain whether race-out-of-place suspects are consistently held to a lower evidentiary standard of suspicion than race-in-place suspects.

Designed by Grogger and Ridgeway (2006), the veil-of-darkness test attempts to address the known deficiencies of earlier approaches for assessing racial disparities, particularly the lack of a reliable benchmark indicating the population at risk of being stopped (legitimately). The veil-of-darkness strategy exploits natural variation in sunlight throughout the year as well as daylight savings time and is based on the premise that officers who are engaged in bias-based policing will be less likely to discern a driver's race when it is dark. If the stops made after the sun is at least six degrees below the horizon have a smaller proportion of Black drivers than the stops made in sunlight, this suggests that "there is racial bias against black drivers" (Grogger and Ridgeway, p. 881).

The major strength of the veil-of-darkness test is that it does not assume that the racial composition of drivers on the road matches that of local residents or that there are no racial differences in traffic violations. Instead, the test's primary assumption is that Black and White motorists do not alter their driving behavior when there is daylight versus darkness (holding constant clock time and season). Importantly, existing research on this assumption suggests that Black drivers relative to White drivers adjust their driving habits such that they are more likely to drive cautiously in daylight (Kalinowski, Ross, & Ross, 2017; Smith et al., 2019). This, along with other aspects of the method (e.g., that it ignores artificial street lighting), means that the veil-of-darkness test is conservatively biased; a null result does not prove the absence of racial discrimination, but uncovering even a modest disparity can indicate a serious problem (Smith et al., 2019).

Following Grogger and Ridgeway's (2006) original design, we limited the stops analyzed to those occurring in the inter-twilight period spanning the earliest and latest times that dusk occurs throughout the year (approximately 5pm to 9pm in the case of Philadelphia). Consistent with existing research in this area (e.g., Pierson et al., 2019), we focus on stop comparisons involving non-Latinx White and non-Latinx Black motorists and we exclude stops occurring during the roughly 30-minute interval between sunset and the end of civil twilight (as this period might be considered neither dark nor light). In line with Grogger and Ridgeway's (2006) original implementation, the veil-of-darkness test takes the form of a logistic regression model where classification as either Black or White is the dependent variable and whether the stop takes place during darkness is the key independent variable.

We also include several control variables in the models. First, clock time (using 6 splines) is held constant to protect against the possibility that the racial composition of drivers varies throughout the early evening hours. Second, we incorporate police district fixed effects to account for the possibility that certain districts may be policed systematically more after dark due to the particular nature of offenses occurring there. Third, we add an indicator variable for the summer months (June, July, or August) to control for the tourist season in Philadelphia, which could alter the racial composition of drivers on the road. Along the same lines, we include indicators for the weeks surrounding two events drawing considerable non-residential populations to Philadelphia: the 2015 Papal Visit and the 2016 Democratic National Convention (DNC). Following Taniguchi et al. (2017), we also disaggregate the veil-of-darkness test by the sex of the driver because their analyses of data for Durham, North Carolina indicated that clear evidence of racial disproportionately was limited to men.

In addition to the veil-of-darkness test, we also utilize other techniques for assessing racial bias that circumvent the thorny issue of external benchmarking. Rather than attempt to estimate expected stop rates by racial group, stop outcome tests analyze the results of reasonable suspicion or probable cause investigations during stops as a way to assess racial parity in the overall quality of those stops. If, for example, officers were disproportionately conducting routine traffic stops of Black drivers as a pretext for searching vehicles for drugs, one might expect to find lower contraband recovery rates for these drivers. That is, one would expect a lower probability of uncovering contraband when decisions to search vehicles are essentially made before there is any meaningful supporting evidence of wrongdoing.

Simoiu, Corbett-Davies, and Goel (2017) offer an important enhancement to traditional outcome tests with their introduction of the threshold test. This test, which utilizes both search rates and contraband hit rates to infer search thresholds, gives race-specific estimates of the average quality of clues needed to initiate a search for illicit goods.⁵ Importantly, the threshold test addresses the problem of infra-marginality often present in traditional contraband hit rate comparisons.⁶ Standard contraband hit rate tests only measure average outcomes, ignoring differences in variability around averages. More specifically, through a hierarchical Bayesian latent variable model, the threshold test takes into account race-specific variances in contraband possession where it may be factually easier to tell the difference between low- and high-risk individuals for one racial group than the other.

Results

Figure 2 presents a simplified example of the underlying logic of the veil-of-darkness test utilizing a portion of the Philadelphia data for illustration purposes. Looking only at district 7 in

Northeast Philadelphia, an area that is approximately 73% non-Latinx White and 9% non-Latinx Black, non-Latinx Black motorists appeared more likely to be pulled over during the 6pm-7pm window when that period had sunlight than when that period was dark. Black people were 16% of the combined population of White and Black detainees when there was darkness but were 23% of that population when there was natural light. Consistent with the possibility of racial bias, this result suggests that the relative likelihood that a Black person will be stopped rises when daylight increases the visibility of driver race.

Table 1 offers a more comprehensive and rigorous assessment of this possibility. For cases citywide that involved motorists classified as either non-Latinx White or Black, a logistic regression model predicting whether a motorist was Black revealed a statistically significant negative coefficient for low visibility in the inter-twilight period with controls for clock time, year, summer season, the 2015 Papal Visit, the 2016 Democratic National Convention, and the police district where the stop occurred. The coefficient for low visibility in Model 1 suggests that overall, the relative odds of a Black motorist being stopped are about 11% less under the “veil of darkness.”

The coefficients for the year indicators in Model 1 suggest that the proportion Black for the combined population of White and Black detainees was rising for the years in the sample (2018 is the omitted reference category). While the coefficients for the summer months dummy variable and the Papal Visit indicator were not statistically significant, the coefficient for the week surrounding the Democratic National Convention was statistically significant and negative, indicating that Black motorists saw their relative likelihood of being pulled over drop during this brief period. Models 2 and 3 disaggregate the citywide results by sex. In line with Taniguchi et al.’s (2017) findings for the City of Durham, the visibility indicator was only significantly related

to the race of the stopped driver for men (who make up the majority of those stopped during the inter-twilight period). This result raises the possibility that if police officers were targeting certain types of individuals in stop decisions, they were doing so through both race and gender lenses in line with the “criminalblackman” stereotype (Russell Brown 2008).

Table 2 presents the application of the veil-of-darkness test to male stops occurring in majority White and majority Black districts (non-Latinx). While the model applied to data from predominantly White police districts indicated that diminished visibility reduced the detention of Black men (Model 1), this effect was also evident in districts that were predominantly Black (Model 2). In other words, regardless of the racial composition of the surrounding area, Black males were more likely to be stopped when sunlight increased the visibility of the driver’s race than were White males.

Conversely, irrespective of spatial context, being seen as White was always an advantage. Indeed, contradicting common narratives about the symmetry and race-neutrality of out-of-place policing, being classified as White appeared to be an even greater benefit in predominantly Black areas (than in predominantly White spaces). Put differently, the larger coefficient exhibited in Model 2 relative to Model 1 implies that, if anything, Black male motorists may be disproportionately targeted to an even worse extent when they are in place than when they are out of place.

Figure 3 displays race-specific contraband recovery rates (hit rates) from searches in predominantly Black and predominantly White police districts. As can be seen, searches of White motorists tend to be more successful in uncovering contraband than searches of Black motorists. This suggests that police officers may require less evidentiary certainty to initiate a search of a Black driver than a White driver. More important for our purposes, Figure 3 also

illustrates how the racial disparity in contraband hit rates is similar regardless of whether the district is predominantly White or predominantly Black. Like the results for the veil-of-darkness test, the relative advantage of being classified as White actually appears greater in areas where White people are out of place and Black people are not. In predominantly Black districts, contraband recovery rates are about 65% higher for White motorists selected to be searched than for Black motorists.

While comparisons of contraband recovery rates provide an intuitive test for racial bias in stops, this methodological strategy suffers from a number of weaknesses. One potentially important limitation is that the approach does not account for infra-marginality. Figure 4 shows the district-specific results for the threshold test recommended by Simoiu, Corbett-Davies, and Goel (2017) as a way to overcome this limitation of traditional contraband hit rate analyses. As is apparent in Figure 4, all districts exhibited a higher search threshold for White motorists compared to Black motorists. Furthermore, the White estimated search threshold was not closer to the Black estimated search threshold in majority Black districts. Indeed, the greatest racial disparity in favor of White motorists was observed in the 14th district—an area that is nearly 80% non-Latinx Black.

We also conducted three robustness tests to help bolster our conclusions. First, in addition to examining district-level variation in race-specific contraband hit rates and thresholds, we also examined vehicle search rates.⁷ Although Black/White disparities were somewhat less than those based on contraband outcomes, all districts displayed higher search rates for Black motorists than White motorists—with the largest relative advantage for White motorists again occurring in the predominantly Black 14th district. Interestingly, this occurred despite the fact that the residential percentage Black and the White search rate for districts were positively

correlated ($r=.22$, similar to what has been reported in previous studies). Overall, our findings imply that while White people in Black districts may experience additional inspection relative to White people in White districts, they do not draw more suspicion than Black people in Black districts.

Second, while our main models attempt to control for spurious seasonal variation in driving and deployment patterns, an alternative approach would be to limit the sample to a brief period around the switch to (or from) daylight savings time (Taniguchi et al., 2017). While this strategy can greatly reduce the statistical power of the veil-of-darkness test, it can also greatly increase our confidence in the test result (when significant) as it is unlikely that driving behavior would change much right before/after the clock moves forward/back one hour. To implement this robustness test we limited our data to the 15-days on either side of the daylight savings time switches in March and November, reducing samples to about 17% of their original size. In line with our main results, Black male motorists were still significantly more likely to be stopped when there was daylight versus darkness. More specifically, estimating Model 2 of Table 2 for this select sample revealed a veil-of-darkness coefficient for majority Black districts of $-.218$ ($p<.01$).

Third, to assuage concerns about aggregation bias, we also applied the veil-of-darkness models in Table 2 to data linked to police service areas ($N=65$). These are smaller, more homogenous places than police districts and are arguably equivalent to the police beat category used in other jurisdictions. While policing policy is more likely to be made at a higher administrative level, utilizing these subunits allowed us to look at areas where residents were almost entirely of one self-identified racial/ethnic category. Focusing the veil-of-darkness model on police service areas that were over 90% non-Latinx Black ($N=62,864$ stops) produced nearly

identical results in direction and magnitude to those reported for stops in majority Black police districts. Black male motorists were still significantly more likely to be stopped when there was daylight as opposed to darkness. Conversely, even in these overwhelmingly Black areas, White male motorists were significantly advantaged when sunlight made drivers' race more discernible.

Conclusion

We use two tests for bias-based policing that do not rely on the very problematic assumption that the population driving through an area closely matches the demographic composition of local residents: (1) the veil-of-darkness and (2) the contraband search threshold. For both tests, we find evidence suggesting bias against Black motorists in stops in Philadelphia. This evidence is not limited to districts where Black people are in the residential minority and thus potentially subject to out-of-place policing. In fact, similar to Levchak's (2017) stop-and-frisk findings for New York City, the apparent bias against Black motorists is most pronounced in areas that are predominantly Black. Consequently, our findings imply that the unequal policing of Black people is an even greater problem when Black drivers are 'in place' than when they are 'out of place' in Philadelphia. This appears to be particularly true when using the veil-of-darkness test as the key measure (applied to males).⁸

As noted earlier, however, the veil-of-darkness test is notoriously a conservative gauge of discrimination as it does not account for factors that would make stops in daylight and darkness appear equivalent despite the presence of racial bias. For example, the original designers of the veil-of-darkness test, Grogger and Ridgeway (2006, p. 884), note that street lighting and "car profiling" reduce the power of the veil-of-darkness test to "reject the null of no racial profiling." For this reason and others, Grogger and Ridgeway (2006) recommended exclusively using the

veil-of-darkness test to assess the *direction* of any bias and cautioned against assertions regarding the comparative *magnitude* of effects. In our study, it is possible that better artificial lighting in predominantly White districts may be attenuating the size of the veil-of-darkness test result. Therefore, it could be that the coefficients for the darkness indicator would be more similar between majority White and majority Black locations if the models accounted for variation in street lights.

Regardless, the consistent direction of the results for both the veil-of-darkness test and the threshold test contradicts the pervasive narrative that White males are exposed to comparable scrutiny in out-of-place contexts. This is important because those who have defended the legitimacy of out-of-place policing have done so by suggesting that there is nothing capricious or unlawful about noticing the out of the ordinary; it is simply a practical aspect of police work. As Withrow (2004, p. 361) notes in his delineation of a theory of contextual attentiveness, “From the beginning of a police officer’s career, he or she is trained to recognize inconsistent patterns of behavior...It seems reasonable therefore that individuals that are ‘different’ would attract the attention and/or suspicion of a police officer.” But, if it is simply “driving while different” that explains the attention given to minorities in majority White places, one would expect to find equality in the degree of disproportionate police response for White people when they are similarly out of place in Black neighborhoods (Withrow, p. 361). Instead, the results imply bias that only runs in one direction—against Black people.

While our analyses consistently uncovered racial disparities in policing outcomes, it is important to note that our research design did not include any direct measures of racial stereotyping and we did not have information on the motivation of the individual officers making stops and conducting searches. Like almost all existing research in this area, the presence or

absence of bias could only be inferred from evidence regarding (in)equality of treatment. Moreover, like previous studies employing the veil-of-darkness and search threshold tests, our methodology has certain limitations. While we were able to include a variety of controls for seasonal fluctuations in driving behavior in our application of the veil-of-darkness test, we were unable to rule out race-specific differences in driving related to the amount of lighting. Similarly, while the search threshold test is arguably a significant improvement over traditional outcome tests, our data did not allow us to disaggregate by search type (e.g., consent-based or probable cause), and this could affect the results. Future studies should consider applying the search threshold test to more detailed data.

We also recommend that future studies focused on evaluating equity in out-of-place policing be structured in a way that corresponds with the *similarly situated individual* framework in discrimination law. This framework requires evidence that certain categories of people are treated differently than others when placed in nearly identical circumstances. Methodologically speaking, this means ensuring that the central comparison is, for example, between how White people in predominantly Black neighborhoods are treated versus how Black people are treated in those same neighborhoods. While there is certainly value for criminological theory in models comparing White treatment in majority Black areas to White treatment in majority White areas, such models cannot directly inform individual discrimination claims since a variety of unmeasured factors likely differ by neighborhood type. If, for instance, predominantly Black areas are perceived as much more crime ridden than they really are, all individuals, regardless of racial background, may face an elevated risk of being interrogated in this context. This would lead to a positive correlation between the percentage of residents that are Black and the rate at which White motorists are stopped or searched. However, this is not convincing evidence of

White drivers looking out of place since Black drivers may have the same or even higher stop/search rates in these places. Our findings demonstrate that while White people in Black neighborhoods may experience additional monitoring relative to White people in White neighborhoods, they do not draw more investigation than Black people in Black neighborhoods, at least in Philadelphia.

Because Philadelphia is one of the nation's most highly segregated large cities, it is possible that racial stereotypes may be especially amplified for Black individuals in majority Black areas. As Sampson and Raudenbush (2004, p. 320) argued in their seminal research on Chicago neighborhoods, "dark skin is an easily observable trait that has become a statistical marker in American society, one imbued with meanings about crime and disorder that stigmatize not only people but also the places in which they are concentrated." Although more comprehensive research is needed, if police officers tend to categorize majority Black neighborhoods as 'bad areas' irrespective of crime rate, blending in may actually serve to heighten suspicion relative to being seen as out of place (Quillian & Pager, 2001). Our results underscore that beyond working to reduce racial disparities in the policing of Black individuals, it is important to independently consider potential biases affecting the policing of Black communities.

Notes

¹ Directive 8.11 of the Philadelphia Police Department (last updated 6-10-14) articulates the department's commitment to unbiased policing and attempts to clarify the acceptable use of race/ethnicity in enforcement actions. This policy notes: "An officer may take into account the reported race or ethnicity of a specific suspect or suspects based on trustworthy, locally relevant information that links a person or persons of a specific race/ethnicity to a particular unlawful incident(s). However, race/ethnicity can never be used as the sole basis for probable cause or reasonable suspicion."

² Besides our emphasis on neighborhood context, we also differ from Pierson et al. (2019) in our use of multiple years, adjustments for seasonal and major event effects, and disaggregation of stops by driver sex.

³ Future research should consider utilizing the veil-of-darkness and search threshold tests to investigate the possibility of bias in stops of Latinx drivers (and whether any evidence of bias is conditioned by the racial and ethnic composition of the surrounding area).

⁴ The majority Black districts in our sample ranged from 54% to 85% Black and the majority White districts ranged from 55% to 83% White.

⁵ There were 41,453 searches over our study period. Unfortunately, the public data we use do not include the officer's stated rationale behind each search. The data do indicate that arrest is an equally rare event (about 2%) for both Black and White motorists, so incident-to-arrest searches

are unlikely to significantly skew our results. Furthermore, we note that Simoiu, Corbett-Davies, and Goel (2017) reported that the inclusion (or exclusion) of consent searches did not meaningfully alter their findings for the search threshold test applied to vehicle stop data for North Carolina.

⁶ Simoiu, Corbett-Davies, and Goel (2017, p. 1194) offer the following helpful example to illustrate the problem of infra-marginality: “To see this, suppose that there are two, easily distinguishable types of white drivers: those who have a 1% chance of carrying contraband, and those who have a 75% chance. Similarly assume that black drivers have either a 1% or 50% chance of carrying contraband. If officers, in a race-neutral manner, search individuals who are at least 10% likely to be carrying contraband, then searches of whites will be successful 75% of the time whereas searches of blacks will be successful only 50% of the time.”

⁷ The racial disproportionality in search rates was strongly positively correlated with racial disproportionality in estimated search thresholds for police districts. Nevertheless, more than one-third of the variation was unshared. Thus, while related, search rates and search thresholds are empirically quite different.

⁸ Worden, McLean and Wheeler (2012, p. 96) convincingly argue that “evidence of racial profiling would be most likely to emerge among officers whose assignments are the most crime focused.” Thus, it is possible that the elevated veil-of-darkness coefficient in majority Black neighborhoods reflects a greater deployment of officers with a crime-suppression mission (as opposed to traffic safety).

References

Arizona v. Johnny Soto Dean (1975). 112 Ariz. 437, 543 P.2d 425. Supreme Court of Arizona.

Bailey v. City of Philadelphia (2011). Settlement Agreement, Class Certification and Consent Decree. No. 10-cv-05952, U.S. District Court for the Eastern District of Pennsylvania.

Carroll, L., & Gonzalez, L. (2014). Out of place: Racial stereotypes and the ecology of frisks and searches following traffic stops. *Journal of Research in Crime and Delinquency*, 51, 559-584.

Close, B.R., & Mason, P.L. (2007). Searching for efficient enforcement: Officer characteristics and racially biased policing. *Review of Law and Economics*, 3, 263-321.

Glover, K. (2007). Police discourse on racial profiling. *Journal of Contemporary Criminal Justice*, 23(3), 239-247.

Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association*, 101(475), 878-887.

Hannon, L. (2019). Neighborhood residence and assessments of racial profiling using census data. *Socius*, 5, 1-9.

Harris, D. (2002). *Profiles in injustice: Why racial profiling cannot work*. New York, NY: The New Press.

Kalinowski, J., Ross, S. L., & Ross, M. B. (2017). Endogenous driving behavior in veil of darkness tests for racial profiling. Retrieved from: <https://ideas.repec.org/p/uct/uconnp/2017-03.html>.

Kennedy, R. (1997). *Race, crime, and the law*. MA: Harvard University Press.

Levchak, P. J. (2017). Do precinct characteristics influence stop-and-frisk in New York City? A multi-level analysis of post-stop outcomes. *Justice Quarterly*, 34, 377-406.

Massey, D., & Denton, N. (1993). *American apartheid*. MA: Harvard University Press.

Meehan, A. J., & Ponder, M. C. (2002). Race and place: The ecology of racial profiling African American motorists. *Justice Quarterly*, 19, 399-430.

Melamed, S. (2019). As Philadelphia aims to curb racial disparities, why are police stops of black drivers skyrocketing? *Philadelphia Inquirer*, October 14th. Retrieved from: <https://www.inquirer.com/news/philadelphia-police-racial-profiling-stop-and-frisk-vehicle-stops-david-rudovsky-mayor-jim-kenney-20191014.html>.

Novak, K.J., & Chamlin, M.B. (2012). Racial threat, suspicion, and police behavior: The impact of race and place in traffic enforcement. *Crime & Delinquency*, 58, 275-300.

Philadelphia Police Department. (2019). Vehicle and pedestrian investigations [Data file]. Retrieved from <https://www.opendataphilly.org/dataset/vehicle-pedestrian-investigations>.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., ... & Goel, S. (2019). A large-scale analysis of racial disparities in police stops across the United States. Retrieved from <https://openpolicing.stanford.edu/publications/>.

Quillian, L., & Pager, D. (2001). Black neighbors, higher crime? The role of racial stereotypes. *American Journal of Sociology*, 107, 717-767.

Rojek, J., Rosenfeld, R. & Decker, S. (2012). Policing race: The racial stratification of searches in police traffic stops. *Criminology*, 50, 993-1024.

Russell Brown, K. (2008). *The Color of crime: Racial hoaxes, white fear, black protectionism, police harassment, and other macroaggressions*. NY: NYU Press.

Sampson, R. J., & Raudenbush, S.W. (2004). Seeing disorder: Neighborhood stigma and the social construction of broken windows. *Social Psychology Quarterly*, 67, 319-342.

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

Smith, M., Tillyer, R., Lloyd, C., & Petrocelli, M. (2019). Benchmarking disparities in police stops: A Comparative application of 2nd and 3rd generation techniques. *Justice Quarterly*, DOI: 10.1080/07418825.2019.1660395.

Taniguchi, T.A., Hendrix, J.A., Levin-Rector, A., Aagaard, B. P., Strom, K.J., & Zimmer, S.A. (2017). Extending the veil of darkness approach: An Examination of racial disproportionality in traffic stops in Durham, NC. *Police Quarterly*, 20(4), 420-448.

Thompson, A.C. (1999). Stopping the usual suspects: Race and the fourth amendment. *New York University Law Review*, 74, 956-1013.

Tillyer, R., Engel, R.S., & Wooldredge, J. (2008). The intersection of racial profiling research and the law. *Journal of Criminal Justice*, 36(2), 138-153.

U.S. Census Bureau. (2019). American community survey 2016 [Data file]. Retrieved from: <https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2016/>.

White, M.D., & Fradella, H.F. (2016). *Stop and frisk: The use and abuse of a controversial policing tactic*. New York: New York University Press.

Withrow, B.L. (2004). Driving while different: A Potential theoretical explanation for race-based policing. *Criminal Justice Policy Review*, 15, 344-364.

Worden, R.E., McLean, S.J., & Wheeler, A.P., (2012). Testing for racial profiling with the veil-of-darkness method. *Police Quarterly* 15, 92-111.

Figure 1. Number of Vehicle Stops by Month in Philadelphia

Figure 2. An Illustrative Example of the Underlying Logic of the Veil-of-Darkness Test

Note: The selected sample is limited to inter-twilight stops involving either non-Latinx White or non-Latinx Black motorists occurring between 6pm and 7pm for the years 2015 through 2018 on local roads in District 7 ($N=908$). District 7 is in Northeast Philadelphia with a residential population that is approximately 73% non-Latinx White and 9% non-Latinx Black.

Figure 3. Racial Disparities in the Productivity of Vehicle Searches

Note: These calculations are based on vehicle searches occurring between 2015 and 2018 in Philadelphia.

Figure 4. Racial Differences in the Estimated Threshold of Suspicion Needed to Trigger a Vehicle Search in Philadelphia Police Districts

Note: Conceptually, the threshold values can be interpreted as the degree of empirical justification needed to motivate a search. Police districts are ordered from left to right by the percentage of residents who are non-Latinx Black (the 17th, 18th, 16th, 35th, 39th, 22nd, 14th, 12th, and 19th districts are majority Black). Thresholds of suspicion are derived from Bayesian analyses that account for infra-marginality in contraband recovery rates from searches (see Pierson et al. 2019, the data are available here: <https://openpolicing.stanford.edu/findings/>).

Table 1. Logistic Regression Models Predicting Whether a Stopped Motorist is Black, Disaggregated by Sex

Variable	Model 1 (Total)	Model 2 (Female)	Model 3 (Male)
Low Visibility Indicator	-.115*** (.015) [.891]	-.051 (.026) [.950]	-.144*** (.018) [.866]
2015 Indicator	-.278*** (.015) [.757]	-.284*** (.027) [.753]	-.280*** (.019) [.756]
2016 Indicator	-.164*** (.015) [.849]	-.209*** (.027) [.811]	-.145*** (.019) [.865]
2017 Indicator	-.084*** (.015) [.919]	-.117*** (.027) [.890]	-.071*** (.018) [.931]
Summer Month Indicator	-.002 (.016) [.998]	.006 (.028) [1.006]	-.008 (.019) [.992]
DNC Convention Indicator	-.213** (.078) [.808]	-.225 (.147) [.799]	-.207* (.093) [.813]
Papal Visit Indicator	-.062 (.070) [.940]	-.044 (.124) [.957]	-.069 (.085) [.993]
Intercept	.790*** (.040)	1.088*** (.069)	.636** (.048)
Number of Vehicle Stops	307,317	96,467	210,850

Note: The sample is limited to inter-twilight stops involving either non-Latinx White or non-Latinx Black motorists on non-highway roads in Philadelphia for the years 2015 through 2018. All models include police district fixed effects and clock time (not shown). Standard errors are in parentheses. Odds ratios are in brackets. Significance tests are two-tailed. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 2. Logistic Regression Models Predicting Whether a Stopped Motorist is a Black Male, Disaggregated by the Racial Composition of the Police District Where the Stop Occurred

Variable	Model 1 (Majority White Areas)	Model 2 (Majority Black Areas)
Low Visibility Indicator	-.138*** (.040) [.871]	-.249*** (.029) [.780]
2015 Indicator	-.279*** (.040) [.757]	-.316*** (.030) [.729]
2016 Indicator	-.077 (.041) [.926]	-.183*** (.029) [.833]
2017 Indicator	-.079 (.041) [.924]	-.080** (.029) [.923]
Summer Month Indicator	.015 (.040) [1.015]	-.007 (.031) [.993]
DNC Convention Indicator	-.201 (.239) [.818]	-.124 (.144) [.883]
Papal Visit Indicator	.062 (.167) [1.064]	-.100 (.128) [.905]
Intercept	.547*** (.083)	3.052*** (.066)
Number of Vehicle Stops	22,318	147,482

Note: The sample is limited to inter-twilight stops involving either non-Latinx White male or non-Latinx Black male motorists on non-highway roads in Philadelphia for the years 2015 through 2018. Both models include police district fixed effects and clock time (not shown). Standard errors are in parentheses. Odds ratios are in brackets. Significance tests are two-tailed.

* $p < .05$, ** $p < .01$, *** $p < .001$