Data for Change

A Statistical Analysis of Police Stops, Searches, Handcuffings, and Arrests in Oakland, Calif., 2013-2014

Stanford SPARQ
Social Psychological Answers to Real-world Questions

Rebecca C. Hetey, Ph.D.
Benoit Monin, Ph.D.
Amrita Maitreyi, B.S.
Jennifer L. Eberhardt, Ph.D.
Data for Change
A Statistical Analysis of Police Stops, Searches, Handcuffings, and Arrests in Oakland, Calif., 2013-2014

Rebecca C. Hetey, Ph.D.
Benoît Monin, Ph.D.
Amrita Maitreyi, B.S.
Jennifer L. Eberhardt, Ph.D.

June 23, 2016
Chapter 1 | STOP DATA OVERVIEW AND DEFINITION OF TERMS

Our task

In May 2014, the City of Oakland contracted with our team of Stanford University researchers to assist the Oakland Police Department (OPD) in complying with a federal order to collect and analyze data on OPD officers’ self-initiated stops\(^1\) of pedestrians and vehicles by race. Our task was to analyze the reports that OPD officers completed after every stop they initiated between April 1, 2013, and April 30, 2014. These reports are called Field Interview/Stop Data Reports (FI/SDR), and the information they contain is called stop data.

We present our independent, detailed, and rigorous assessment of these stop data in the current document, Data for Change: A Statistical Analysis of Police Stops, Searches, Handcuffings, and Arrests in Oakland, Calif., 2013-2014. In addition, we summarize the findings of this stop data analysis, discuss four other research initiatives, and list 50 recommendations for reform in a second document, Strategies for Change: Research Initiatives and Recommendations to Improve Police-Community Relations in Oakland, Calif.

Our approach

Analysts usually take one of two approaches to police stop data. The first approach is to lay out the evidence for racial disparities in stops, and then conclude that the police are racists who are deliberately targeting people of color. This approach intends to shake law enforcement agencies into changing their ways. Instead, it usually incites so much police resistance that meaningful reform becomes difficult, if not impossible.

The second approach is the opposite of the first: Analysts find no evidence for racial disparities in stops. These analysts often use bloated statistical models so chock-full of covariates (i.e., control variables) that any evidence of disparate treatment disappears. For instance, their reports often conclude that African Americans are more likely to commit crime than are other groups, and so police are just going where the crime is. Everything is as it should be. There is nothing to see here. Yet the daily experiences of communities of color suggest otherwise, and their frustration with these null-finding reports harms relations with police.

---

\(^1\) For a stop to be included in this data set, an officer must have been required to complete a Field Interview/Stop Data Report (FI/SDR). In other words, the stop must have been self-initiated and have involved one or more members of the community who were detained, arrested, or subjected to a search or the request to be searched. Casual encounters in which officers talked to community members, but the community member remained free to leave at any time, are not captured here.
In our stop data analysis, we take a third approach—a problem-solving approach—that concludes with neither attack nor denial. We report some real and significant racial disparities in OPD stops, searches, handcuffings, and arrests, even after accounting for crime rates, demographics, and other factors that influence policing activity. The OPD acknowledges these disparities and is eager to address them. To this end, we have conducted our analyses in a manner that allows the OPD to make evidence-based changes in their policies, practices, and procedures. For example, using statistical models, we have isolated the conditions under which racial disparities are greatest and least. Simply knowing where, when, and how racial disparities are likely to emerge gives the agency direction on how to lessen them. This approach has yielded dozens of tactics that the OPD and other law enforcement agencies can undertake to reduce racial disparities. In other words, our approach both acknowledges existing racial disparities in policing and gives police the tools they need to mitigate and perhaps even eliminate these disparities.

Soon, a new California assembly bill (AB 953) will require law enforcement agencies across the state to collect the sort of stop data we have analyzed here. Yet to date, many law enforcement agencies are not sure how to use their data to make change, as they lack a common model for addressing racial disparities in a productive way. Here we offer a model of how policing agencies can use data to solve problems, instead of using data to attack or deny.

As researchers, we can apply a problem-solving approach only when law enforcement agencies value, trust, and understand this approach. The OPD is such an agency. The OPD leadership has given us unprecedented access to the data on which our work relies. They understand that our findings may be unfavorable at times, yet they are poised to address any racial disparities that come to light. Because of their progressive position, we now understand more about improving police-community relations than ever before. On this issue, the OPD has contributed greatly to the Oakland community, many other communities, and the law enforcement industry as a whole.

Overview of the data

During this 13-month time period, 28,119 stops were recorded by 510 sworn OPD officers. Each of these officers made an average of 55 stops during the 13-month period under examination (median number of stops = 35, Interquartile Range \(^2\) = \([9, 82]\)). It is thus worth keeping in mind throughout this report that the median number of documented stops for a given officer was only one stop every

---

\(^2\) The Interquartile Range (IQR) is a summary of a distribution based on dividing a data set into quartiles, or four equal parts. The IQR lies between the first (25%) and the third (75%) quartiles—it thus literally describes the values occupied by the 50% of observations located in the middle of the distribution. The IQR is a useful measure of central tendency as well as of the variability in a sample or population that is not too affected by outliers, and does not make implicit assumptions about the shape of this distribution (for example, compared to the standard deviation):
Chapter 2 | METHODOLOGY

In this stop data report, our aim is to understand whether or not race influenced the rate at which people of different racial groups were stopped by the Oakland Police Department and whether or not race affected the course of a given stop. The first question could take the form of “Were African Americans more likely to be stopped because of their race?” The second question focuses on what we will call post-stop outcomes. More specifically, once the stop was made, was the subject of the stop more or less likely to be handcuffed, searched, and/or arrested as a function of his or her race? We focus on these three post-stop outcomes because they are of interest to many stakeholders.

Members of the community often complain about being handcuffed at high rates, even in cases in which the stop does not ultimately end in an arrest. Similarly, many members of the public are concerned about the frequency of searches being conducted, especially when the majority of searches overall (more than 70%) lead to no recovery. The OPD in particular is interested in what it can do as an organization to improve search recovery rates. As for arrests, our goal was to examine what is arguably the most severe outcome of a stop.

A police stop is a complex interaction between two people: the officer and the community member. Each individual brings something to the interaction, but is also, to some extent, bound by the other person’s behavior. Each person is continuously acting and reacting to the other person. A police stop, then, can be thought of as a complex and dynamic system in which each actor reacts in real time to cues emitted by the other actor in the interaction. It thus can be difficult to confidently assign causality when analyzing stop outcomes. To take a few examples, if we observed a systematic effect of officer gender, it could be a mistake to attribute it only to officer behavior. Imagine we found, for example, that female officers handcuffed community members less often. The more obvious interpretation is that for whatever reason female officers may be hesitant or reluctant to use handcuffs and therefore they use handcuffs less often. It could also be the case, however, that maybe female officers appear less threatening to community members, which causes the community members to not behave in ways that can lead to the kind of escalation that may require handcuffing. Thus, in this example, it is not actually the officer’s gender but the reaction in the civilian that is elicited by the officer’s gender that triggered the need for handcuffing. Similar problems of causality emerge in the study of parenting, such as in cases in which some children’s traits can trigger certain behaviors in parents, which in turn are causal in shaping children’s development. Similarly, when we look for examples of quality of life indices predicting stop outcomes it is important to take into account that these environmental factors may affect both actors going into the interaction, and that both then react to each other’s initial attitude, creating complex dynamics that can spiral out of control without being directly attributable to either actor.

On the surface, the question of how race shapes policing decisions may seem fairly straightforward. It seems plausible enough that one could examine raw counts of the number of stops that were made
of each racial group and compare them to each other. Similarly, one could examine the number of times each post-stop outcome occurred as a function of the race of the person stopped. Stop data reports produced by the Oakland Police Department typically examine this type of raw data.\(^{37}\) Such reports are produced as part of the OPD Stop Data program, which is designed to promote transparency and allow “the Department to assess effectiveness and identify potentially biased behaviors.”\(^{38}\) This method, clearly, is considered to have some merit.

The question of whether or not race plays a causal role in policing decisions, however, is quickly complicated by the fact that we cannot know for certain what the officer was thinking when he or she made the decision to pull someone over or conduct a search. Even if the stop data form had a question that read, “Did you decide to make the stop because of the community member’s race?” this problem would still not be solved. There is a strong prescriptive norm, or societal consensus or demand, that people should be egalitarian and should not express racial prejudice or discriminate by using race as the basis of their decisions.\(^{39}\) The fear of being seen as a racist, and even of the disciplinary or legal consequences of profiling, would likely compel most officers, to simply answer “no” to this question regardless of whether race did play a factor in their decisions.\(^{40}\) Furthermore, psychologists have found that, in many cases, people are surprisingly inaccurate at knowing and

\(^{37}\) Stop data reports produced and published by the Oakland Police Department are available at this link: [http://www2.oaklandnet.com/government/o/OPD/a/PublicReports/index.htm#stop](http://www2.oaklandnet.com/government/o/OPD/a/PublicReports/index.htm#stop)


\(^{40}\) Indeed, social psychological research has shown that there are a variety of reasons individuals might want to respond without prejudice. For some people, the motivation to respond without prejudice is internal: they avoid acting in prejudiced ways and relying on stereotypes because doing so would violate their own personal values and concept of themselves as egalitarian. For others, the motivation to respond without prejudice is external: they try to appear nonprejudiced in order to avoid the negative reactions and disapproval of others. See Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*(3), 811-832. Further, another powerful motivating force is to fit in with others and to conform to social norms about when it is and is not acceptable to express prejudice and discriminate and about which groups it is and is not acceptable to express prejudice against. See Crandall, C. S., Eshleman, A., & O’Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359-378. To avoid appearing racist, people will go out of their way to avoid mentioning or acknowledging race, even when this avoidance is counterproductive. See Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology, 95*(4), 918-932.
articulating the reasons for their behavior, judgments, and decisions.\textsuperscript{41} Especially in the case of race, bias is often not explicit. Rather, much of racial bias tends to be \textit{implicit}, or a bias that people are not even aware that they have.\textsuperscript{42} Implicit bias can come from repeated exposure throughout one's lifetime to subtle cultural cues that are transmitted to us, for example, when we watch television, interact with our parents, or more generally observe the stereotypical ways in which different groups are commonly depicted, treated, and talked or thought about.\textsuperscript{43}

Thus, to the extent that a given officer's true motivations and attitudes are \textit{unknowable}, we argue that posing the question of whether or not particular officers are “biased” is not the most fruitful way to begin an investigation of how race may influence police stops. Indeed, this question of individual-level bias can be counterproductive and something of a nonstarter. Even in the absence of biased or racist individuals, institutions themselves can be biased by having policies and structures


in place that harm some people and favor others, even in unintended and unanticipated ways. Institutional disparities can perpetuate themselves: Simply being exposed to evidence of inequality can cause people to become more supportive of the very policies that produce that inequality. It can be tempting to look for a window into the hearts and minds of individuals (e.g., police officers), but when researchers and practitioners focus instead on institutions (e.g., the criminal justice system, a specific law enforcement agency), they are in a better position to measure and evaluate the consequences of policies and practices and identify whether some groups disproportionately bear the burden of any negative outcomes. In the academic, legal, and policy arenas (and beyond), this disproportionate burden is referred to as disparate outcomes or disparate impacts. The hunt for bias can also unfortunately lead to name calling and defensiveness, and can thus become counterproductive. Here, we acknowledge that the issue of individual officer bias is inherently unanswerable. As researchers, we can never know what was inside a particular police officer’s head. Instead, in this report, we focus on and examine what we can know: whether or not there are systematic differences in outcomes of stops for different groups, controlling for as many factors as possible that could legitimately justify such differences; and whether or not police officers’ decisions to make stops and to handcuff, search, and arrest have disparate impacts on people of color in the community.

What is a benchmark?

Of all OPD stops that were made from April 1, 2013, through April 30, 2014, 60% were of African Americans. Stops of African Americans were made at a rate of more than three times that of Hispanics, the next most common racial group that was stopped. Sixty percent may sound like a large percentage; however, we cannot know if 60% is high or low or begin to understand the role of race until we have some figure to which to compare this number, or some larger context in which to consider it. What should the number be? At what rate would we expect African Americans in the City of Oakland to be stopped? To begin to answer this question, we need some point of reference, or benchmark. Thought of another way, we cannot know whether race played a significant role until we have first accounted for other factors that might plausibly explain why we would expect to

---


see racial differences. As a salient example of such other factors, let us consider crime rate. When a claim of racial profiling is made against a police department, perhaps the most common rebuttal is that the police are simply going where the crime is, presumably acting under the theory that increased police presence reduces crime; there is evidence that police patrol does, in fact, deter crime. To the extent that high-crime areas have higher concentrations of African Americans and to the extent that it is an effective policing strategy to stop more people in these high-crime areas, then that means that, indirectly, more African Americans will be stopped. In this scenario, race is not the driving factor; crime is. To the extent that race and crime actually predict each other, one might observe apparent racial differences in stop rates that in actuality have nothing to do with race at all and are really accounted for by crime rate in the neighborhood.

There is no consensus in the academic literature about which benchmarks or other factors are the most appropriate to take into account. In fact, selecting benchmarks to include in statistical

---

46 Ayres & Borowsky, 2008; Analysis Group, Inc. (2006), Pedestrian and motor vehicle post-stop data analysis report, prepared for City of Los Angeles.
49 There is no established procedure for how to select benchmarks. Some researchers begin with population data and compare the rates of stops made of African Americans to the African American population share (usually measured by census data; e.g., Bailey, et al., Plaintiffs v. City of Philadelphia, et al., Defendants, *Plaintiffs’ fifth report to court and monitor on stop and frisk practices*(2015), C.A. No. 10-5952, In the United States District Court for the Eastern District of Pennsylvania; New York Civil Liberties Union (2012), *Stop-and-frisk 2012: NYCLU briefing*). Other researchers, particularly those analyzing vehicle stops, focus instead on measures of driving behavior, such as traffic data about involvement in motor vehicle collisions (e.g., Lovrich, et al., 2007). Still other research teams, such as Analysis Group (2006) in its treatment of the LAPD’s stop data, include more than a dozen control variables that include characteristics of the officer who made the stop, in addition to multiple measures of crime rate, and other variables about the location of the stop, including the neighborhood’s economic well-being and stability (Analysis Group, Inc. 2006, Pedestrian and motor vehicle post-stop data analysis report, prepared for City of Los Angeles). There is also no established procedure for what to do with benchmarks in an analysis once they have been selected. Some researchers include benchmark variables as control variables in large statistical models. The Analysis Group (2006) in its treatment of the LAPD data on vehicle and pedestrian stops did this. Other researchers adopt a criterion for specifying what they will consider a significant race difference. For example, in an analysis of data resulting from vehicle stops made by the Washington State Patrol, Lovrich, Gaffney, Mosher, Pratt, and Pickerill (2007) decided that a racial difference would not be considered “substantively significant as long as the percentage of those contacted in any particular racial group is not more than five percentage points” greater than the percentage of the group in the benchmark comparison” (p. 5; original emphasis; Lovrich, et al., 2007).
analyses can be in and of itself a fairly contentious issue.\(^50\) In the next section, we will review which benchmarks have been most commonly used in analyses of stop data and policing decisions. As you will see, no benchmark is perfect and each comes with its own set of strengths and weaknesses.

**What benchmarks have been used in past research?**

*Population demographics:*

Suppose we were analyzing the stop data for the police department of a major American city and found that 75% of all stops were of African Americans. With this information alone, it is impossible to tell whether or not this police department might be engaging in policing practices that disproportionally affect the African American community. As one benchmark, we might want to know something about the local population. *Whom the police stop is necessarily limited to the universe of people the police could stop.* Should we learn that our hypothetical police department were in Ann Arbor, Michigan, which according to the 2010 Census is only 8% African American, then we might conclude that the police department was stopping African Americans at a very high rate given the demographics of the residents. If, however, the police department were a few miles away in Detroit, Michigan, which is 83% African American, then the rate at which the police department was stopping African Americans would more closely mirror the demographics of the residents. If anything, we might conclude that the police department was stopping fewer African Americans than we would expect given the population demographics.

As in this example, one place many researchers start is with an examination of local population demographics. Usually researchers acquire relevant census data about what share of the total population is made up of those who fall into different demographic categories. By breaking down stops and census information at the level of census tract (the geographical subdivision used by the Census, typically between 1,000 and 8,000 inhabitants), researchers can control for demographic factors as they vary within a city. The racial demographics of an area is usually the central focus, but other types of information are collected as well to the extent that these other demographic variables are also suspected of changing the likelihood at which people are at risk of being stopped. For instance, the police stop younger people more than they stop older people. Unemployed people might similarly be stopped more often to the extent that they might be out driving during the day when many people are at work. Researchers usually take into account or compare how many people

---

\(^{50}\) As we have said and will discuss in more detail, many researchers rely on census data about the racial demographics of the neighborhoods where stops were made to establish a sense of who may be at risk of being stopped by police in the first place. Grogger and Ridgeway (2006) are critical of the use of such data when trying to estimate risk for being pulled over in a vehicle. To get around the need for any external benchmarks, they instead created an approach to test for possible race differences that they call the “veil of darkness.” Another controversy surrounds what the appropriate use of benchmarks is.
are stopped by race to the share of that racial group in the population.\textsuperscript{51} Indeed, an analysis done by the New York Civil Liberties Union (NYCLU) of the New York Police Department’s (NYPD) stop-and-frisk rates in 2012\textsuperscript{52} appealed to population demographics. In the data highlights section of the NYCLU report, the authors noted that:

Young black and Latino men were the targets of a hugely disproportionate number of stops. Though they account for only 4.7\% of the city’s population, black and Latino males between the ages of 14 and 24 accounted for 40.6\% of stops in 2012. The number of stops of young black men neared the entire city population of young black men (133,119 as compared to 158,406).\textsuperscript{53}

The NYCLU report further used demographic data to take on the argument that so many stops are of African American and Latino community members because they happen to live in high-crime precincts that are predominately African American and Latino. The report noted that even in most of the 10 precincts with the lowest percentages of African American and Latino residents in the entire city (comprising between 8\% and 14\% of the population) more than 70\% of stops were of African Americans and Latinos.\textsuperscript{54}

Though comparing the rates at which certain racial groups are stopped to the rates at which those groups are present in the general population makes intuitive sense, this approach has limitations. First, in some cases, census data may systematically undercount undocumented residents and migrant workers, an issue that has been noted as a significant problem when trying to obtain accurate information about the percentage of Hispanics who reside in a given area.\textsuperscript{55} Second, most of the data on racial demographics include all residents of a particular area, regardless of their age or other characteristics. A particular census tract might be 50\% African American, for example, but a significant portion of those African American residents might be small children or the elderly, who are statistically less likely to be stopped by police compared to 18- to 30-year-olds. To the extent that racial minority groups tend to have higher birth rates, resulting in a population that skews younger than majority groups,\textsuperscript{56} using Census data that include all residents regardless of age would make for


\textsuperscript{52}New York Civil Liberties Union (2012), Stop-and-frisk 2012: NYCLU briefing.

\textsuperscript{53}New York Civil Liberties Union (2012), p. 2.

\textsuperscript{54}New York Civil Liberties Union, p. 6.

\textsuperscript{55}Nicholas P. Lovrich et al. (2007), Results of the monitoring of WSP traffic stops for biased policing: Analysis of WSP stop, citation, search and use of force data and results of the use of observational studies for denominator assessment, Report to the Washington State Patrol relating to: National Highway Traffic Safety Administration (NHTSA) Grant-Funded Study on Racial Profiling Phenomena in Washington State OGRD # 107828.

a conservative test of the role of race in police decision-making. Relatedly, population demographics often do not take into account how many residents have driver’s licenses or otherwise drive regularly, which is of particular importance in areas in which the majority of police stops are vehicle stops. Driving behavior, then, may be another important variable, which we will return to shortly. Another limitation of population demographics is that people routinely venture away from where they live, e.g., to go to work, school, or church or to go shopping. People are not always stopped where they live.\textsuperscript{57} Heavily commercial areas, for example, might not even have a sizeable population of residents, but they do attract people who inhabit those spaces and therefore can be stopped there. People are not supposed to be stopped by police merely because they are physically present somewhere, but rather because they are suspected of breaking some law. Even if most people in a given city are of one race, if that group commits traffic violations or commits crimes at low rates, then that group should not be stopped often. For this reason, the utility of population demographics may be overestimated. Overall, the major limitation of relying on population information is that the demographics of residents may not reflect the demographics of those who actually could be stopped.

\textit{Crime rate:}

Another common approach is to rely on data about crime. When using crime rate as a benchmark, researchers often include the rates of violent crime, property crime, or both.\textsuperscript{58} To the extent that self-initiated police stops are part of a larger enforcement strategy to prevent crime, the distribution of stops throughout a city should mirror the distribution of crime. In fact, members of the law enforcement community often say that they are simply “going where the crime is.” Therefore, to the extent that self-initiated stops and other policing strategies are concentrated in high-crime areas that happen to be predominately African American,\textsuperscript{59} then it makes sense that African Americans will, indirectly, be stopped more frequently overall because they are more likely to be physically present in areas where a larger proportion of stops are made. One limitation to using crime rate data as a benchmark, of course, is that simply being in a high-crime area does not, on its own, provide justification to make a stop. Indeed, living, working, or otherwise traveling in a high-crime area


\textsuperscript{57} For more on this point, see Ridgeway, G. (2009). Cincinnati Police Department traffic stops: Applying RAND’s framework to analyze racial disparities. Santa Monica, CA: RAND Corporation.


\textsuperscript{59} Past research has shown that high crime, on the one hand, and segregation and poverty in African American communities, on the other hand, tend to go together. Racial segregation concentrates poverty, which, in turn, concentrates crime and violence. Massey, D. S. (1995). Getting away with murder: Segregation and violent crime in urban America. \textit{University of Pennsylvania Law Review, 143}(5), 1203-1232.
does not, on its own, predict one’s likelihood of being directly involved in criminal activity.\textsuperscript{60} In the case of \textit{Brown v. Texas}, the U.S. Supreme Court ruled that it is unconstitutional to stop someone based solely on the fact that he or she is in a high-crime area.\textsuperscript{61} Note also that crime rates may not be as accurate as one might think. Quillian and Pager, both sociologists, explain that crime rate data, which usually comprise crimes reported to police by the public, may systematically undercount the true numbers of crimes because a fear or mistrust of the police may keep members of the community from filing reports.\textsuperscript{62} Other reasons people may not report crime to police: They feel that less serious crimes are too trivial to report, the police would not be interested or would not be able to do anything in response, and they can privately deal with the incident themselves.\textsuperscript{63} Quillian and Pager argue that people’s perceptions of crime are colored by the racial demographics of the neighborhood. The researchers found that even after controlling for a host of factors including two measures of crime (the crime rate collected by police and a separate measure based on people’s self-reported victimization of crime), the percentage of young Black men\textsuperscript{64} living in the neighborhood predicted residents’ views of how much crime there was in the neighborhood. It is an open empirical question how susceptible police officers might be to the influence of stereotypes equating African American men and crime in determining what they consider to be a high-crime neighborhood and, more consequently, how they make policing decisions when in those areas.

What about the argument that a difference in crime rates by race could justify racial disparities in stops? Consider a federal court case regarding the New York City Police Department’s (NYPD) stop

\begin{footnotesize}

\footnote{Brown v. Texas (443 U.S. 47 1979). Note that in Illinois v. Wardlow (528 U.S. 119, 2000), the U.S. Supreme Court considered the character of the neighborhood to be a legitimate factor in finding reasonable suspicion to stop someone, although the Court ruled that it cannot be the sole justification for a stop. The Court allowed that an officer needs only two factors to establish reasonable suspicion under the Fourth Amendment: being in a high-crime area and unprovoked flight from police. Ferguson and Bernache (2008) describe how the same behavior, apparently running from police, means two different things depending on whether the neighborhood is considered to be a “low-crime” or “high-crime” area. In a low-crime area, the police do not automatically have justification for making a stop, whereas in a high-crime area, the police do. The authors argue that this difference has implications for residents of high-crime areas: “‘High-crime areas’ are a fact of constitutional law: individuals in those areas have different Fourth Amendment protections than they would in other locations in the same town, city, or state.’ (p. 1589). See Ferguson, A. G., & Bernache, D. (2008). The ‘high-crime area’ question: Requiring verifiable and quantifiable evidence for fourth amendment reasonable suspicion analysis. \textit{American University Law Review}, 57, 1587-1644.}


\footnote{Tarling, R., & Morris, K. (2010). Reporting crime to the police. \textit{British Journal of Criminology}, 50(3), 474-490.}

\footnote{Defined as being between the ages of 12 and 29 years old. Note the researchers also controlled for the total percentage of young men (of all races) who fell into this age range, so their effects cannot be dismissed as being driven by the presence of young men more generally.}
\end{footnotesize}
and frisk program that received a great deal of attention. In 2013, a federal judge ruled that stop, question, and frisk tactics were unconstitutional because the NYPD violated the rights of racial minorities by subjecting them to high numbers of stops and searches.66 During the trial, New York Police Commissioner Raymond Kelly did a television interview with the ABC program “Nightline,” during which he defended the legitimacy of these policies by appealing to the crime rate by race:

About 70% to 75% of the people described as committing violent crimes—assault, robbery, shootings, grand larceny—are described as being African-American... The percentage of people who are stopped is 53% African-American, so really, African-Americans are being under-stopped in relation to the percentage of people being described as being the perpetrators of violent crime.67

Commissioner Kelly argued that because African Americans commit more violent crimes, then it follows that African Americans should be stopped more often, presumably on the suspicion that they are more likely to be violent criminals. Legal scholar David Cole explains that, indeed, many criminologists have concluded that Blacks, men, and younger people do in fact commit crime at a higher per capita rate than Whites, women, and older people. “Thus, all other things being equal, it is rational to be more suspicious of a young black man than an elderly white woman. But that it may be rational does not make it right.”68 Cole describes how the correlation between race and crime remains a stereotype to which most African Americans do not conform. Even if African Americans are actually more criminal in their behavior, he reasons, only about 2% of African Americans are arrested each year for committing any crime. Thus, the vast majority of African Americans are not charged with crimes. A police department that relies exclusively on race in making the decision to stop people, then, is likely to stop many more innocent people than guilty people. In addition to being less than effective, using “an individual’s race as a direct proxy for that individual’s criminality is legally problematic under current prohibitions against racial profiling.”69 In the State of California, California Penal Code Section 13519.4(e) prohibits racial profiling by law enforcement. According to the OPD’s Departmental General Order M-19, racial profiling can be defined as:

---

67 But note that, according to the 2010 Census, the population of New York City is 25.5% Black or African American alone, meaning people who identify as one race (Black or African American) and no other race (this figure does not take into account Hispanic origin). Therefore, African Americans were overrepresented among those stopped relative to their general share of the population.
69 Ayres, I., & Borowsky, J. (2008), A study of racially disparate outcomes in the Los Angeles Police Department, Prepared for the ACLU of Southern California, p. 4.
The use of race, ethnicity, or national origin in determining reasonable suspicion, probable cause or the focus or scope of any police action that directly or indirectly imposes on the freedoms or free movement of any person, unless the use of race, ethnicity, or national origin is used as part of a specific suspect description.  

**Driving behavior:**

In many jurisdictions, the majority of stops are vehicle stops that are made because of traffic violations. Consider, for example, various state highway patrol agencies that are responsible for enforcing traffic laws on a large stretch of interstate highways and freeways. Since the primary purpose of state highway patrol agencies is to provide traffic enforcement and keep roads safe, then crime rates are less relevant. Highway and freeway drivers may also be more likely to live far away from where they are stopped. In fact, they may not even live in the same state. Population demographics, then, may also be of less use. Accordingly, researchers who conducted an analysis of traffic stops made by the Washington State Patrol argued that driving behavior is the more appropriate benchmark. Their measures of driving behavior included contacts initiated as a result of calls for service and vehicle assists, as well as contacts initiated as a result of radar patrols (e.g., drivers who were identified as speeding via radar and thus the Trooper was "blind" to the identity of the driver), and collision data. In particular, the researchers argue that collision data coded by race is the most effective benchmark of driver quantity and quality that provides “a reliable and cost-effective indicator of driver population demographics.”

The problem is that many agencies do not collect this information or have the resources to store and make such information searchable or user-friendly. In addition, because many traffic stops are made due to alleged equipment failure (e.g., broken taillights), it is unclear to what extent actual driving behavior (e.g., speeding, reckless driving) would provide an accurate base rate for the likelihood of being pulled over. The rates at which members of different racial groups are stopped and cited due to equipment failure might have more to do with the year, make, and model of the cars they tend to drive than with how they drive. Additionally, people who tend to drive older vehicles, and/or

---

70 Oakland Police, Departmental General Order M-19 (November 15, 2004), p. 1
71 See, for example, Missouri Revised Statutes, (August 28, 2015), Primary purpose of highway patrol, Chapter 43, Highway Patrol, State, Section 43.025.1, which states "the primary purpose of highway patrol is to enforce the traffic laws and promote safety upon the highways" and California Highway Patrol, (2016), which states that the CHP was created to "provide uniform traffic law enforcement throughout the state" and that "assuring the safe, convenient and efficient transportation of people and goods on our highway system is still our primary purpose." Retrieved from https://www.chp.ca.gov/home/about-us
74 Lovrich, et al., (2007), p. 2; see also p. 12 for details on collision data.
people who are unable to afford maintenance of said vehicle (e.g., unable to afford the timely replacement of a headlight or taillight) may be more likely to be stopped for equipment violations. Socioeconomic status, then, may be a useful benchmark. Another possibility is that getting pulled over is driven, in part, by how attention-grabbing a vehicle is. People have a commonsense notion that certain types of cars lead to more traffic tickets. For instance, a Google search of "red cars get more tickets" yields more than 29 million results. Perhaps different racial groups, and certainly different socioeconomic groups, are more or less likely to drive "flashy" or attention-grabbing cars. Another issue is that vehicle stops may sometimes be made in service of law enforcement purposes other than the enforcement of traffic laws. Pretext stops are legal and "objectively valid" traffic stops, that is, stops based on genuine traffic infractions, wherein a separate motivation for the stop is to "search for evidence of an unrelated offense." In these so-called "pretext" stops, an officer's goal may be to gather intelligence or engage in some other enforcement strategy and the actual traffic violation is the legal justification to make the stop. Although we cannot know whether a traffic violation stop is or is not a pretext stop, driving behavior may not always be a useful benchmark if such stops do occur. Therefore, in all of these cases, it is debatable whether driving behavior is an informative benchmark to use.

**Internal benchmarks:**
Rather than obtaining external data (e.g., census data) with which to compare an agency's stop data, some researchers have used internal benchmarks that they constructed from within the stop data itself. In an analysis of Cincinnati Police Department traffic stops done on behalf of the RAND Corporation, Ridgeway (2009) constructed an internal benchmark for each officer:

This method selects an officer, identifies stops that other officers made at the same time and in the same neighborhoods, and compares the racial distributions of the stopped drivers. Since the officers are patrolling the same area at the same times, the racial distributions should be the same (assuming that the officers are on the same assignment).

This basic approach has been adopted in many cities as part of Risk Management and "early-warning systems" designed to identify problem officers. The OPD uses this approach, as does the

---


76 In Whren v. United States, 517 U.S. 806 (1996), the Supreme Court ruled that if and when police officers have probable cause for a traffic stop, a pretextual motive does not invalidate the stop. They held that "the temporary detention of a motorist upon probable cause to believe that he has violated the traffic laws does not violate the Fourth Amendment’s prohibition against unreasonable seizures, even if a reasonable officer would not have stopped the motorist absent some additional law enforcement objective (pp. 809-819)."

When it comes to an analysis of stop data, however, different officers may be on different assignments, especially in smaller police forces with limited personnel. We might expect that officers working different assignments would have different patterns of stopping people by race. This method also relies on researchers having access to accurate data about what assignment each officer was working during any given stop throughout the course of the entire time period from which the stop data have been collected.

**Time of day:**

Another benchmarking method that has been used to test for the existence of racial bias in policing decisions is known as the veil-of-darkness method. As described by Ridgeway (2009), this method compares the breakdown of stops by race made during the day to the breakdown of stops by race made at night. The basic logic is as follows: If a police department were targeting African American drivers to stop, evidence of this practice should be most apparent during daylight when the race of a driver is presumably most visible to the officer. If race is not discernible at night, then police are simply less able to racially profile at night. Ridgeway cautions that an “overly simplistic implementation of this analysis” would simply compare the percentage of African American drivers stopped during the day to the percentage of African American drivers stopped at night. A number of researchers do simply take into account whether a stop was made during the day or at night, although they usually do so alongside other benchmarks and control variables. This is considered “overly simplistic” because other variables are likely to vary as a function of time of day. For instance, if African Americans were simply less likely to drive at night than during the day relative to members of other racial groups, then we would expect a lower percentage of African Americans to be stopped at night, but this would prove nothing about suspected racial profiling. To combat this problem of variables that are linked or confounded, Ridgeway proposes using Daylight Saving Time as a natural experiment. The pattern of stops by race can be observed on one Monday, when it is still light out at 6:30 PM, and can be compared to the pattern of stops by race on the following Monday, when it is dark out at 6:30 PM (because clocks have been set back one hour). Everything

---


else has been held constant, except for the visibility of race. This type of analysis tends to only include stops made during the inter-twilight hours, usually from approximately 6:00 PM to 8:00 PM, because they could be made during daylight one week and at night the next week. This approach, then, significantly reduces the number of data points, which may reduce the statistical power to such an extent as to even preclude the possibility of detecting statistically significant racial bias. This limitation may or may not be acceptable to interested parties whose goal is to make policy recommendations based on the results of a given analysis.

A reliance on time of day as a benchmark also tends to work better for vehicle stops than for pedestrian or bicycle stops, in which the officer is more likely to be physically close to the subject of the stop—thereby making race more apparent—when the decision to make the stop is made. The subject of the stop is likely to be in plain sight as opposed to being enclosed and occluded by a large, usually moving, vehicle. Another limitation of time of day as a benchmark is that it is debatable to what extent the race of a driver is truly obscured at night and to what extent it is clearly visible during the day. Especially on city streets, as opposed to a dimly lit highway or freeway, streetlights may make it entirely possible to see the race of the driver at night, especially when an officer has the opportunity to first closely follow a car. Relatedly, many vehicles have tinted windshields and windows that can make it impossible to determine the race of the driver, even during broad daylight. Another limitation to the argument that nighttime lighting conditions completely blind an officer to race is that the make and model of the car itself, or any number of other factors (e.g., demographic makeup of the location) may act as a proxy for the race of the driver.

**What role do benchmarks play in statistical models?**

*“Controlling for” vs. interactions: Understanding the role of additional variables*

Benchmarks are a class of factors that theoretically should be of importance in establishing a baseline, or base rate, of the outcome in question. In this research, the outcomes of interest are the rates at which members of different racial groups in Oakland are stopped, handcuffed, searched, and/or arrested by the OPD. When building statistical models to test for the existence of significant racial disproportionality, we need to translate our theoretical benchmarks into concrete, operationalized, and measurable variables. For example, to include a crime benchmark in our model, we could count the number of times a given criminal offense was committed within a neighborhood (or census tract), and divide that by the number of residents in that neighborhood. For instance, if we count the number of murders that took place and divide that number by the number of residents in that same location, we have a concrete murder rate to serve as our crime variable. We can then include that variable in our model. When we “take into account” or “control for” the effect of a variable, we call that variable a “covariate.”

---

As a growing number of law enforcement officials have recognized, there is a pressing need for more and better data on policing to guide policymaking, promote transparency and accountability, and identify and address the instances of bias where they arise. Throughout its Final Report, the President’s Task Force on 21st Century Policing urged law enforcement agencies to collect and make public data on a variety of department practices, including all police-citizen encounters such as stops, frisks, searches, summons, and arrests.

Yet, as the Task Force articulated, it is essential that law enforcement agencies be given the resources and guidance necessary to collect, analyze, and report their data in meaningful ways. Stop data in particular poses a number of challenges, from establishing uniform standards on the types of encounters that ought to be reported, to developing the appropriate benchmarks against which to compare demographic variation, to providing the context necessary to interpret the results properly.

Consistent with the Task Force recommendations, the California state legislature recently passed a bill requiring all law enforcement agencies in the state to collect and report to the California Attorney General’s (CAAG) Office demographic and other data on all stops conducted by officers in the state. The law instructs CAAG to develop guidelines and technical specifications to guide law enforcement agencies in complying with the new requirements.

We propose to partner with the Center for Policing Equity and the Policing Project at New York University School of Law (“Research Partners”) to develop a set of best practices for the collection, analysis, and reporting of stop data—and ultimately to develop a guidebook for use by law enforcement agencies both in California and nationwide.
Need for Guidance on Data Collection, Analysis, and Reporting

In designing and implementing a stop data program, law enforcement agencies—and oftentimes, state and local legislatures—must resolve a number of difficult questions around the collection, analysis, and reporting of the data.

Collection

One of the threshold questions that all jurisdictions must answer is what data to collect—and in what form. In addition to basic demographic data—such as race, gender, and age—departments must decide whether to collect data on location; time of day; the subject’s perceived mental or physical disability, English proficiency, and housing status or homelessness. (Even the most basic demographic data poses the question of whether to ask the subject or rely upon officer perception.) Departments also must decide how much information to collect about the encounter itself, including the officer’s basis for suspicion, whether a search was conducted, and its outcome—as well as whether to collect this data in structured or narrative form.

Each of these choices presents a variety of tradeoffs and considerations, including officer time, individual privacy, and the effect that data collection might have on the encounter itself. For example, including a narrative description of an encounter poses obstacles to quantification or systematic analysis, is time consuming for patrol officers, and requires significant investment in training in order to produce uniform reports. However, forms that include only check boxes, such as NYPD’s much-scrutinized UF 250 may promote a culture of minimal effort that is counterproductive both to identifying patterns of criminal behavior and increasing officer accountability.

Indeed, another critical question—and seemingly so basic that jurisdictions all too often overlook it—is how to define a “stop” for data collection purposes. Under the Fourth
Amendment, a person is “stopped” or “seized” when a reasonable person in his or her situation would not feel free to leave or otherwise terminate the encounter. But if the goal is to track encounters that may reveal evidence of bias, or contribute to a lack of trust between law enforcement agencies and their communities, this definition may be under-inclusive. Such a definition would not capture what is sometimes referred to as a “mere inquiry” (e.g. approaching someone on the street to ask where they are going, whether they live in the area, or even to request consent to search). In addition, it is possible that officers and community members—particularly minority residents in heavily-policed communities—may have very different perceptions of whether the individual stopped is in fact free to leave. Officers may need additional guidance—or a different definition entirely—to ensure that they record all relevant encounters.

Data collection also presents a number of practical and technological challenges that departments will need to anticipate in designing their programs. Departments across the country are at various stages in their abilities to capture, analyze, and report on behavioral data. Some departments use paper forms to track police behavior, while others hand enter information from paper forms into computer systems. Some have computer systems that are unable to generate the necessary reports to conduct data analyses. An audit of this process, and how to move data collection forward so it is captured, stored, and analyzed, is much needed in the policing community. Additionally, reviewing what technologies are available to meet the needs of departments with a range of capacities will be a significant benefit.

*Analysis*

Perhaps the greatest challenge facing departments is analyzing and making sense of the data they collect. Here we highlight just a few of the analytical difficulties that departments face.
One of the key objectives in collecting stop data is to identify—and ultimately address—the existence of bias in policing. Yet there is at present no consensus on how best to measure biased police enforcement—and several methodological challenges to doing so.

Much of the difficulty in measuring bias results from two distinct problems: the “Numerator Problem” and the “Denominator Problem.” The “numerator” in this case is the number of people of each race or gender who are stopped by the police. The problem is that there are no uniform standards for the collection of stop data across departments (and sometimes internally within a single department), which makes it difficult to know for sure how many people are actually stopped, and to compare stop and hit rates across jurisdictions. The “denominator” is the benchmark population against which stop and hit rates must be compared to know whether minorities are disproportionately affected. Given the demographic variation in crime rates, as well as differences in day-time and night-time populations, or automobile use rates, most agree that total population is not a proper benchmark. The problem is that there is no consensus in the field as to what the benchmarks should be, and even some of the most robust techniques (such as economic equilibrium modeling of stop outcomes or “hit rates) can produce conflicting results and interpretations from the same data (e.g., Dahrmapala & Ross, 2004; Knowles, Persico, & Todd, 2001; Sanga, 2009).

In addition to the methodological obstacles to evaluating stop data, there also are potential legal and policy considerations that departments must take into account. For example, state laws differ with respect to the legal consequences that attach to the fact that a government policy or practice disproportionately affects racial minorities (or other protected classes). Under some state’s laws, disproportionate impact is enough to trigger legal scrutiny; in other states there must be evidence of discriminatory intent. Departments will need to keep these background
legal considerations in mind in deciding both how to interpret their data and whether or how to adjust their practices in response to what their analyses show. Departments also may wish to consider the effect of any perceived disparities on stops data—even if entirely lawful—on community trust, and would benefit from guidance on how to distinguish between those disparities that are unavoidable and those that could be mitigated through training or policy change.

Finally the issue of potential bias is made all the more complex by problems of context, which bear upon collection and analysis both. Law enforcement officials often recognize that data indicates disproportionate demographic impact. Yet, how to evaluate this impact may differ depending on the causes. For example, are the stops officer-initiated, or the result of calls for service? A slate of these different contextual variables must be developed, and then implemented at both the collection and analytic stages.

**Reporting**

Departments also need guidance on how to report data to their communities so as to promote transparency and accountability. Agencies will need to decide whether to release incident-level data in addition to aggregate reports—and if so, will need guidance on how to ensure that data are fully de-identified to protect the privacy of those involved. Also, as discussed above, departments will need to find ways to present evidence in context and to educate the public so that community members understand what the data actually show, and what conclusions the data support.

**Existing Resources and Literature**

Although there are a number of resources that deal with specific aspects of stop data collection, none offer the comprehensive guidance that departments need in order to implement
their programs. Much of the existing literature is focused on finding appropriate benchmarks against which to judge stop data. For example, Lori Fridell’s seminal *By The Numbers: A Guide for Analyzing Race Data from Vehicle Stops* (2004) provides a detailed overview of various approaches to analyzing traffic stop data for evidence of bias—and makes some preliminary recommendations on how to collect and categorize stop data (see also Grogger & Ridgeway, 2006). The report, however, does not provide the practical steps that departments must take to implement such recommendations without the help from outside researchers. And while it identifies multiple methods for data analyses, departments are tasked with distinguishing the best metrics for their agency, something that social science researchers have yet to find consensus on. Greg Ridgeway’s *Analysis of Racial Disparities in the New York City Police Department’s Stop, Question, and Frisk Practices* (2007) provides additional guidance on the benchmarking problem in the context of pedestrian stops. (See also Walker 2001).

A far smaller number of resources focus on collection and reporting. A 2003 report issued by the COPS office, McMahon et al., *How to Correctly Collect and Analyze Racial Profiling Data* (2003), offered some preliminary thoughts on these topics, but made clear that additional research and guidance is needed for departments looking to implement data collection programs.

Our proposed project will build on existing research in two important ways. First, it will build on existing benchmarking studies by giving departments the tools necessary not only to identify bias where it exists, but also to determine whether bias stems from department policy, individual officer training or biases, or some combination of both. By leveraging CPE’s Justice Database—which includes standardized data from over fifty law enforcement agencies on stops, department policies, and officer surveys—we will highlight patterns in stop data that may be
indicative of each of the potential root causes of biased enforcement.

Second, our guidebook will offer much-needed guidance on data collection, integration, and reporting, which are topics that existing resources address only preliminarily or in passing. In particular, the guidebook will offer guidance on how best to define a stop, and train officers on the definition, to ensure accurate and comprehensive reporting; how to choose among existing technologies to support data collection efforts; and how to integrate stop data into existing record management and early warning systems. We also will develop a template (or set of templates) for reporting stop data in ways that are informative and accessible to the community, and which present sufficient context for the public to understand what the data actually show.

**Implementing California Assembly Bill 953**

These general observations are consistent with what the California Attorney General’s Office has found in the course of its early efforts to implement Assembly Bill 953 (“AB 953”), the Racial Identity and Profiling Act of 2015, which went into effect this year and will eventually require all California law enforcement agencies to collect and report stop data to the CAAG. Large agencies—those employing over 1,000 sworn officers—must submit their first reports by April 2019. Smaller law enforcement agencies—those with fewer than 334 officers—have until April 2023 to comply.

Although AB 953 sets out a number of minimum reporting requirements—such as the time and location for the stop, as well as the perceived race, gender, and approximate age of the person detained—it instructs the Attorney General’s Office to promulgate additional guidelines to define what constitutes a stop, instruct agencies on what additional data to collect, and to develop technical specifications for collection and reporting to ensure data are standardized across jurisdictions.
Beginning to collect stop data is an enormous undertaking that may have profound implications both for the public and for officers on the front lines—and it therefore is imperative to get the implementation right. At present, many agencies do not collect any stop data, and the Oakland and Berkeley police departments are the only two law enforcement agencies in the state that routinely publish stop data. Thus, over four hundred additional law enforcement agencies will need guidance on how to resolve all of the various questions discussed above.

Although the Attorney General’s Office has begun to work through some of these issues and to involve stakeholders in the process, it is clear to us that we will need to draw heavily on the experience and expertise of social science methodologists and legal professionals who have been deeply involved in working on these issues. On the legal side, we have been grappling with a variety of questions, from “what exactly constitutes a stop,” to how to account for the presence of passengers, and whether and how to supplement an officer’s initial perception of the subject’s demographics with actual demographic information. On the technical side, it is imperative that we bring the best thinking to bear on this challenge. As reporting burdens grow, we run the risk of unintentionally shifting law enforcement behavior, so it is very important that we seek ways to collect this data without taking officers away from their primary job of keeping communities safe. It also is essential to determine the appropriate benchmark population against which to assess the data we collect—and to account for such things as driving patterns or differences between daytime and nighttime populations.

Proposed Project

The California Attorney General’s Office will partner with the Center for Policing Equity and the Policing Project at New York University School of Law (“Research Partners”) to develop a set of best practices on data collection, analysis, and reporting—and to produce a
guidebook to assist law enforcement agencies in California and throughout the United States.

We also will work closely with two local law enforcement agencies—in Oakland and Richmond—to learn about their data collection practices and capabilities, test out our recommendations, and learn from their experience in formulating the guidebook. For this pilot, we selected one large and one mid-sized agency to ensure that our guidebook addresses the very different challenges that large and small or mid-sized agencies are likely to face during the implementation process. We also made sure to select agencies that are at different stages in the implementation process so as to both have the benefit of experience from a department that is further along, and an opportunity to pilot the implementation with a department that is just starting out.

- **Oakland Police Department:** The Oakland Police Department (OPD) is a large agency with 741 sworn officers serving a community of 400,000. The OPD was among the first agencies to join the White House Police Data Initiative, and is one of two agencies in California that currently collects and makes public data on all police-citizen encounters. Every six months, the OPD issues through its open data portal a report with detailed breakdowns of its stop data along demographic and other dimensions. Working with Oakland will enable us to learn from the department’s current practices in formulating our initial recommendations. At the same time, OPD has expressed great interest in working with our Research Partners to assist the department in fine-tuning its practices to comply with the statute and best practices.

- **Richmond Police Department:** The Richmond Police Department (RPD) is a mid-size law enforcement agency with 178 officers in a city of approximately 100,000. The RPD also is a participant in the White House Data Initiative, although at the
moment the RPD only makes public its data on the use of force. The RPD currently
does not collect any stop-level data. The department is in the early stages of
developing a stop data collection program, but its initial roll-out will include only a
portion of the data that Richmond ultimately will be required to report under AB 953.
Although the RPD has until 2023 to begin reporting stop data to the Attorney
General’s Office, the department has agreed as part of this project to participate in a
pilot to test our preliminary guidelines and to get the benefit of our Research Partners’
expertise in designing and implementing its data collection system.

In addition to working with these specific agencies, this Project will draw upon other
agencies’ learning experiences and expertise. Given that the CAAG is the primary law
enforcement applicant, we will be able to learn from the experiences of many California agencies
that are at the very beginning of implementation under AB 953. In addition, we have at hand
CPE’s experience with several other California jurisdictions who have begun to collect stop data
voluntarily—but have yet to make this public. As part of CPE’s National Justice Database,
departments already have provided researchers access to their data on pedestrian stops, vehicle
stops, and use of force. Access to this broader database should provide a clearer indication of
how to resolve various analytic questions (e.g., “the denominator problem”) by isolating the
relative contribution of individual officers, departments, and exogenous factors (e.g., residential
demographics, crime, economic indicators, etc.)

**Strategy to Achieve Program Outcome and Goals**

To achieve the above stated goals, we propose the following project plan. We will work
closely with—and draw on the expertise of—our Academic Partners in carrying out all of the
project objectives and deliverables.
The Psychological Science of Racial Bias and Policing

Jillian K. Swencionis
John Jay College of Criminal Justice

and

Phillip Atiba Goff
John Jay College of Criminal Justice

© American Psychological Association, 2017. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at:

http://content.apa.org/doi/10.1037/law0000130

Please direct correspondence to:

Jillian K. Swencionis
Center for Policing Equity
John Jay College of Criminal Justice
524 West 59th Street
New York, NY 10019
jill@policingequity.org
Abstract

What can the social psychology of racial bias teach us about the potential for racial bias in policing? Because social psychological research is mostly laboratory based and rarely includes police officers, direct generalizability is limited. However, social psychology has identified robust risk factors that make individuals more likely to engage in disparate treatment—even without overt prejudice. This article maps these situational risk factors to common experiences in modern patrol policing. Specifically, we identify the following situations common to patrol policing as risk factors that make bias more likely to result in discrimination: discretion, novice status, crime focus, cognitive demand, and identity threats. Where possible, we also review studies that include officers, and take place in policing-relevant contexts. With the map provided in this article, we exhort psychologists to translate previous laboratory findings to field settings to advance the practice of democratic policing and expand the science of bias.

(149 words)
Introduction

In the two-and-a-half years since Officer Darren Wilson shot and killed Michael Brown, Jr. in Ferguson, Missouri, violent encounters between police and communities have frequently dominated America’s news cycle. The fact that so many of these encounters feature officers shooting a Black man—who is often unarmed—has also heightened concerns that race plays a troubling role in policing and in the broader society. For instance, a representative poll of U.S. adults revealed that 60% perceived fatal police encounters with Black Americans to be signs of a broader problem of racism (Morin, Parker, Stepler, & Mercer, 2017). As a result, the public, researchers, and policymakers continue to ask: What is the role of racial bias in policing? This is a question psychological science should be uniquely well suited to address.

Social psychology in particular boasts a voluminous literature on the mechanisms responsible for discrimination (Fiske, Gilbert, & Lindzey, 2010). Of note, the scientific consensus is that, in addition to personological, individual difference factors that dominate lay conceptions of racial discrimination (e.g., trait-based prejudices, authoritarianism, and social dominance orientation), situations play an important role in predicting discriminatory behaviors (Dovidio, 2001; Goff, 2013; Goff, Steele, & Davies, 2008; LaPiere, 1934). Consequently, social psychology conceptualizes of discrimination (as it does for nearly all behavior, see Epstein & O’Brien, 1985; Mischel, 1968) as a combination of personal risk factors modified by situational ones. In the context of policing, therefore, social psychology should be able to identify the elements of policing that are most likely to produce racially discriminatory outcomes—possibly resulting in the loss of liberty or life.

But it does not. At least, not yet.
Despite a sizable literature and an appropriate framework for diagnosing contexts within policing that are vulnerable to discriminatory behavior, few empirical psychological studies engage police officers or even realistic police-relevant contexts. As a result, questions about the role of racial bias in policing do not have a robust scientific literature from which to find answers. The goal of this article is to build a theoretical bridge between two worlds: the experiences of patrol officers and the psychological science of racial bias. By aligning known predictors of discrimination with the situational risk factors present in modern policing, we hope to outline both a set of risk factors for bias in policing that can be useful to practitioners, and a research agenda for scholars who care about how basic psychological science translates to this vital domain. While personal and situational risk factors are both important, in this article we focus on situations because they suggest especially actionable possibilities for intervention.

**A Framework for Understanding Racial Bias in Policing**

We organize the article in terms of five experiences common to contemporary policing, each of which corresponds to robust categories of risk factors for discriminatory behavior: discretion, being a novice, crime focus, cognitive demand, and identity threats. Each of these features of modern patrol policing have one or more psychological risk factors associated with them (see Table 1). In each section, we detail the psychological risks most associated with the experience (referring back to previous sections where a psychological risk factor applies to more than one patrol experience). Along the way, the article speaks to gaps and inconsistencies in the psychological research that bear on these domains and are especially critical to questions about police policy and practice.

It is important to note as we explore risk factors for biased behavior in policing, this article does not engage deeply with the character of police, which might be described by various
individual differences. Among the most robust findings in psychology is that biased behavior need not be a reflection of individual character. Additionally, situations present opportunities for reducing the influence of bias on behavior, as opposed to individuals’ levels of bias, which are likely more resistant to change. For these reasons, the article is framed in terms of situational risk factors rather than personality types.

We have chosen to focus on the experiences of patrol officers in part because the majority of contact between law enforcement and residents occurs with officers patrolling a “beat.” To be sure, non-patrol activities also present risks for engaging in discriminatory behavior. For example biases may skew how/when/whether an officer interviews witnesses, structures a lineup, or interprets evidence. But these are beyond the scope of the present article.

Table 1

Map of Policing Risk Factors to Psychological Routes to Discrimination

<table>
<thead>
<tr>
<th>Psychological Routes to Discrimination</th>
<th>Situational Risk Factors</th>
<th>Discretion</th>
<th>Novice Status</th>
<th>Crime Focus</th>
<th>Cognitive Demand</th>
<th>Identity Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDO</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aversive Racism</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereotypicality</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Shooter” bias and Black-crime associations</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Bias in nonverbal behaviors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Stereotype threat</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Masculinity Threat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instead, this article focuses on identifying the research on racial bias that has included actual patrol officers as participants and discusses the research most relevant to policing contexts that has been conducted with non-officers, largely in laboratory settings. In most of these cases,
new research is needed to determine how generalizable laboratory research is to actual policing contexts. We begin with the experience of discretion.

**Discretion**

Police officers have more decision-making latitude about what happens to community members the lower they are in the organizational hierarchy, with officers on patrol duty given a great deal of discretion (Walker & Katz, 2011; Wilson, 1980). Officers on patrol face situations in which there is no one mandated course of action, and so they are free to make decisions throughout an interaction with a civilian, any of which can have substantial impact on the communities they are sworn to protect. At an even more basic level, discretion includes not just decisions within an interaction, but whether an interaction happens in the first place, and with whom. Understanding the role that situational ambiguity plays in officer decision making, much of police training is geared towards helping officers make sense of ambiguous situations (Fielding, 1988; Haberfeld, 2002). Still, there is no training that can prepare officers for the broad array of contexts they will encounter. And this chronic exposure to ambiguity overlaps with the psychological literature on bias, which consistently observes that ambiguity leads to a higher likelihood that of discriminatory behavior. What follows is a review of the literature that identifies ambiguous situations as a robust risk factor for engaging in discriminatory behaviors.

**Discretion and Social Dominance Orientation**

Situations in which officers can exercise discretion can make disparate outcomes more likely because discretion allows officers’ prejudice to influence their decisions (Dovidio, 2001; Kalev, Dobbin, & Kelly, 2006). One form of explicit prejudice that may be more likely to influence decisions in patrol situations when officers may exercise discretion, is Social Dominance Orientation (SDO). SDO measures individuals’ support for hierarchies that
discriminate against members of lower social strata (Sidanius & Pratto, 1999). This literature demonstrates that individuals who self-report higher levels of SDO are more likely to engage in discriminatory behavior when given the opportunity. While many forms of explicit prejudice have declined over time, SDO continues to exert an influence on behavior without the expression of this attitude having been diminished by prevailing social norms (as demonstrated by numerous surveys and experiments).

SDO tends to be higher in police officers compared to members of the general public, college students, and public defenders, even controlling for demographic variables (Sidanius, Liu, Shaw, & Pratto, 1994). In this combined laboratory and field study, Sidanius and colleagues recruited Los Angeles Police Department officers, public defenders, students, and members of the general public recruited from jury pools, and showed individuals with higher levels of SDO tended to choose to go into policing (i.e. they have pre-existing higher levels of SDO). Given this evidence of higher SDO in policing and the ability for discretion to permit individual biases to influence behavior, the link between SDO and officer decision-making should be a high priority among researchers.

**Discretion and Aversive Racism**

Having more discretion, and in particular having multiple possible explanations for one’s behavior, increases the likelihood that discrimination will result from a form of bias known as aversive racism. In the literature on aversive racism, individuals explicitly affirm their egalitarian values, but on some level they also hold implicit biases and/or negative affective responses to members of stigmatized groups, a dissonant feeling they find uncomfortable or “aversive” (Dovidio & Gaertner, 2004). Recognizing this contradiction between one’s values and actions can pose a threat to a person’s self concept, heightening their concern about appearing racist. The
literature on aversive racism is largely laboratory-based, and no studies to the current authors’
knowledge have investigated aversive racism in policing. In student and general samples,
though, aversive racists are at higher risk of engaging in discriminatory behavior in situations in
which the social norms or individuals’ accountability for their actions are ambiguous.

Laboratory experiments of aversive racism demonstrate how ambiguous social situations
with multiple possible courses of action and alternative explanations for discrimination can
present opportunities for aversive racism to translate into discriminatory patterns of behavior.
For instance, in one study, White women college students were randomly assigned to hear a call
for help from either a Black or a White person (Gaertner, Dovidio, & Johnson, 1982). Some
participants were assigned to hear the call for distress while engaging with a group (who were
non-responsive to the call for help) while others were alone when they heard the call.
Participants who were alone responded equally quickly to Black and White calls for help.
However, participants heard the distress call while engaged with a (non-responsive) group were
nearly twice as fast to respond to someone who was White and in distress than someone who was
Black. Consistent with the aversive racism formulation, when group norms are not strong (i.e.,
you are in a group that does not respond to calls for help), racial bias emerges.

In fact, this particular combination of low explicit bias and high implicit bias predicts a
range of discriminatory behaviors, from interpersonal interactions to selection decisions. For
instance, in another study, college undergraduates engaged in either an intra-racial or interracial
interaction. During the interaction, participants’ explicit attitudes predicted differences in their
verbal behavior toward Black versus White interaction partners. On the other hand (Dovidio &
Gaertner, 2000; Gaertner & Dovidio, 1986), their implicit attitudes, measured by reaction times
to pairings of Black and White faces with positive and negative words, predicted differences in
their nonverbal behavior toward their interaction partner (Dovidio, Kawakami, & Gaertner, 2002). Importantly, the interaction partners rated participants’ warmth and friendliness as corresponding more closely to participants’ non-verbal behaviors, than to participants’ verbal behaviors, producing a mismatch of participants’ self-perceptions of friendliness, and their partners’ perceptions of their friendliness.

Aversive racism also predicts discrimination in outcomes like hiring and college admission decisions (Dovidio & Gaertner, 2000; Gaertner & Dovidio, 1986). Though a preference may not emerge between highly qualified Black and White applicants, when choosing between middling applicants, aversive racism results in a strong preference for the White applicant versus the Black applicant. Though conducted in the laboratory, these studies serve as proof of concept that aversive racism results in discriminatory behaviors and biased outcomes, especially when accountability for participants’ actions is ambiguous – in this case, when choosing between middling applicants, participants could easily attribute their choice of the White applicant to some factor other than applicant race.

Importantly, while no research has directly examined aversive racism in policing, it is reasonable to hypothesize that increased levels of officer discretion may also facilitate biased behavior in line with an aversive racism framework. Similarly, because standards of probable cause and reasonable suspicion—important elements in determining officer behaviors towards community members—are notoriously difficult to specify, increased police/civilian contact with an enforcement mindset may also produce the kind of normative ambiguity that promotes racially disparate treatment. Consequently, it appears that policing is a fertile domain in which to pursue future research in aversive racism, and we encourage researchers to take up the opportunity.
Discretion and Stereotypicality

The degree to which someone looks like the stereotype of one’s group may also influence perceptions, decision-making, and behaviors in ambiguous situations. For instance, laboratory research by Kahn and colleagues demonstrated that stereotypicality influences standard metrics of implicit bias (Kahn & Davies, 2011) and are also associated with racial disparities in field settings (Kahn, Goff, Lee, & Motamed, 2016). In a random sample of booking photos, Kahn and colleagues found that, controlling for type of arrest, reported level of resistance, and the presence of drugs and alcohol in a suspect’s system, ratings of suspects’ phenotypic stereotypicality were negatively associated with severity and likelihood of force for White suspects. That is, the more stereotypically “White” a suspect looked, the less force was used on them even when controlling for other factors. Interestingly there was no relationship between stereotypicality and force for Black suspects in the same sample.

The stereotypicality of a target has also been linked to higher levels of “shooter bias” in computer simulations (Kahn & Davies, 2011), and even an increased likelihood of the death penalty (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Again, this research suggests that further research on the role of suspects’ racial stereotypicality in police decisions is both necessary and urgent.

Being a Novice

Police officers receive many different forms of training, beginning before they interact with civilians in an officer capacity. General training tends to focus on the specific kinds of tasks involved in patrol policing, providing the opportunity to practice and develop skills. Separate trainings tend to focus on reducing bias. Even in general trainings, though, mere practice and repetition tend to help officers to complete tasks more accurately, reducing the influence of bias.
on their behavior even if bias is not specifically addressed. In general, training is ongoing, practice and expertise are crucial, and in the meantime officers will encounter situations for which they have not have had the opportunity to prepare.

As individual officers gain experience, they are vulnerable to the effects of being a novice. For officers, this may manifest as they acclimate to their patrol responsibilities, to police culture, or to the cultures of the neighborhoods they patrol. Across several domains in psychology, the literature on task expertise concludes that individuals with more practice tend to be more likely to complete tasks accurately (e.g., MacLeod, 1998), and less likely to engage in discrimination (e.g., Kawakami, Dovidio, & van Kamp, 2005; Plant & Peruche, 2005). Researchers have built on this literature, showing various domains in which practice helps to build expertise, and further, how practice on accuracy in weapon identification and shooting decisions may reduce racial biases. In the domain closest to policing, researchers have examined the effects of practice—or training—in the context of laboratory shooting simulations. These tasks are thought to reflect bias that stems from participants’ implicit associations of crime-related concepts with Black men, as demonstrated in one study by statistical correlations between student participants’ awareness of Black-crime stereotypes, and their racial bias in the decision to “shoot” images of unarmed Black versus White individuals in a first-person shooting task (Correll, Urland, & Ito, 2006).

**Novice Status in “Shooting” Situations**

Payne and colleagues created one of the first laboratory “shooting” related tasks, observing racial bias in student participants’ errors when detecting whether they saw a gun or a tool, and showing empirical support for one reason why shooting decisions tend to be racially biased: people may be more likely to mistake an object for a gun, when primed with images of
Black compared to White men (Payne, 2001; for more detail, see the section on crime focus). Plant, Peruche, and Butz (2005) created a similar task, but during each of 160 experimental trials, gave participants a choice to shoot or not shoot a target person who was accompanied on-screen by either a gun or other object. Over the course of these trials, student participants made racially biased shooting errors at first, for example being more likely to “shoot” unarmed Black targets compared to unarmed White targets. By the second half of the task, these biased errors decreased to non-significant levels. In a controlled experiment, Plant and colleagues gave some participants their race-weapon task, and other participants a control task – one day later, the participants who practiced the control task showed racial bias on the race-weapon task, but the participants who had practiced the race-weapon task did not. This experiment therefore demonstrated the causal role of a training task in reducing the expression of racial bias in shooting errors. Importantly, Plant and Peruche (2005) replicated the effect of practice on diminished racial bias in shooting errors in a sample of 50 police officers.

Correll and colleagues created another variant of this first-person shooter task, showing participants a series of experimental trials each with an image of either a Black man or White man holding a gun or object in different naturalistic settings, and presenting a choice to either “shoot” or “don’t shoot” within a short deadline on each trial, as short as 630 milliseconds (Correll, Park, Judd, & Wittenbrink, 2002). Correll and colleagues have shown that both students and officers in laboratory simulation settings show racial bias in shooting errors, erroneously “shooting” unarmed Black targets more frequently than unarmed White targets, and doing so faster, and also choosing “don’t shoot” more frequently in response to armed White targets compared to armed Black targets, and also doing so faster. Correll and colleagues also

---

1 A meta-analysis supports the robustness of these effects (Mekawi & Bresin, 2015).
brought in police officers and community members to participate in their shooter task, and observed that community members showed racial bias in both their error rates and reaction times, while police officers showed bias only in their reaction times, and not in their error rates (Correll et al., 2007). A subsequent study of undergraduates in the laboratory demonstrated one possible reason why participants with more expertise tend to exhibit higher accuracy: when these undergraduates practiced the shooter task and then returned two days later, their accuracy tended to increase and bias tended to decrease (Correll et al., 2007). However, the accuracy-improving effects of training have also been shown to be diminished when Black-crime stereotypes are relatively more salient, showing how a crime focus can attenuate this beneficial effect of practice (Sim, Correll, & Sadler, 2013).

These studies begin to reveal how practice and expertise can reduce racial bias in the decision to shoot, among students, community members, and officers—at least in simulated laboratory settings. Further research is needed to address the role of training, practice, and expertise in officers’ actual decisions, or at the very least, in settings that more closely approximate actual police work. These questions may become more relevant in regard to modern policing settings where novice officers have contact with civilians in higher-crime areas, where reducing crime is especially salient—and thus crime-related stereotypes may be especially accessible. These decisions, too, are likely to be impacted by high levels of officer discretion, in which more officer decisions may be subject to bias (and may be improved by more practice). Importantly, the “shooter” simulations described in this section all institute short response deadlines, which create time pressure that is relevant to real-life situations in which a person may be armed. In contrast, some studies that have attempted to determine the role of racial bias in officer decisions to shoot have allowed participants unlimited time in which to make their
decision – a feature that makes these decisions less similar (and perhaps less relevant) to potentially dangerous interactions in real life, and allows participants unlimited time in which they may correct for the effects of bias on their decisions (e.g., James, Vila, & Daratha, 2013). Further research will need to take notice of this critical aspect of officer decision-making. The next section discusses time pressure in more detail, as one example of a situational determinant that makes discrimination more likely.

**Novice Status and Explicit Prejudice**

Returning to a form of explicit prejudice described earlier in the section on discretion, SDO is also related to officers’ levels of experience, in that more experience actually predicts higher levels of this explicit form of prejudice. In addition to the selection effect described earlier, in which officers have higher levels of SDO than members of the general population, an earlier, longitudinal study of officers over the course of an 18-month training found that SDO additionally increases the longer an officer serves on the force (Teahan, 1975). This study suggests that SDO, which broadly predicts greater levels of racially discriminatory behavior (Sidanius & Pratto, 1999), is both more common in police recruits and increases further with time on the force. While existing findings do not provide sufficient grounds for making stronger claims, these findings provide sufficient reason to call for additional research in this area.

**Crime Focus**

Policing has always emphasized officers’ role in controlling crime, even at the expense of other important goals such as protecting the public safety, providing service to the public, having positive interactions with citizens, or being accurate. In contemporary contexts, crime focus might include assigning more officers to patrol higher-crime neighborhoods in pursuit of a crime deterrent effect. Modern incentive structures reinforce the focus on crime control, often
rewarding arrests and citations but often not helpfulness or accuracy. This chronic focus on crime control can also act as a risk factor for discrimination because it heightens the accessibility of pervasive stereotypes associating Black people (in general) and Black men (in particular) with crime-related concepts (Eberhardt, Goff, Purdie, & Davies, 2004). We have already seen that novice status tends to increase the effects of some of these Black-crime stereotypes, and next we will review how crime focus can act as its own risk factor.

Showing effects of crime focus on stereotype-based mis-identification of suspects by real officers, Eberhardt and colleagues (2004) recruited police officers to participate in a study where participants were presented with a picture of a potential suspect. The suspect was either Black or White. Immediately before seeing the suspect, participants were either subliminally primed with words associated with crime (e.g., crime, investigate, arrest, apprehend) or nonsense letter strings. After officers saw the suspect, they were given a surprise memory task that involved picking the original suspect out of a lineup of other suspects of the same race. Each of the five faces in each lineup were pretested so that they could be arrayed in a scale from most to least phenotypically stereotypical for their racial group. The target face was the mid-point of that stereotypicality scale (with two more stereotypical faces and two less stereotypical faces also displayed). Results revealed that officers who had been primed with crime-related words were no more or less accurate than those who had not. However, when they misremembered (which was the majority of the time) they made more stereotypical errors for Black suspects when primed with crime than when unprimed. Likewise, they made less stereotypical errors for White suspects when primed with crime than when unprimed.

Eberhardt and colleagues (2004) specifically manipulated crime focus, which influenced the extent to which racial stereotypes influenced their [mis-]identification behavior. In situations
like this, discretion (discussed above) and cognitive demand (discussed below) are also relevant risk factors. In particular, this finding recalls the effects of discretion on the risk of making errors due to a person’s stereotypicality (Kahn & Davies, 2011; Kahn et al., 2016), errors that can additionally be amplified by a situational focus on crime.

As another prominent example of racial stereotypes strengthened by a focus on crime, Payne and colleagues have developed a program of research showing a bias toward identifying guns (compared to non-gun objects) following brief exposure to Black faces compared to White faces (Bishara & Payne, 2009; Govorun & Payne, 2006; Payne, 2001, 2005, 2006; Payne, Lambert, & Jacoby, 2002; Payne, Shimizu, & Jacoby, 2005). In particular, this research has shown that participants are more likely to mistakenly identify non-gun objects as guns when the objects are paired with images of Black men’s faces, compared to when they are paired with images of White men’s faces. In these experiments, participants receive instructions to identify each object that appears on a computer screen as either a gun or a tool—while ignoring images of faces.

In addition to the roles of novice status and crime focus as risk factors that increase the influence of Black-crime stereotypes, these stereotypes also pose an especially high risk in situations in which officers have to make decisions very quickly, and/or in which there are multiple demands on their cognitive capacity—both also common features of modern patrol policing. We discuss these risk factors next.

**Cognitive Demand**

One reason modern patrol policing is such a challenging job is that officers experience sometimes-life-threatening situations in which they must make a decision very quickly, or in which there are many simultaneous demands on their attention. Psychological models of human
attention tend to show that people have limited capacity for processing information simultaneously (e.g., Broadbent, 1958; Navon & Gopher, 1979; Pashler, 1994), and there is no reason to believe policing would be an exception. The literature on cognitive demand and racial bias holds that when the demands of a situation exceed the processing resources available, the processing resources that are typically devoted to controlling the expression of racial bias are otherwise occupied, and are more likely to fail. As noted in the section on crime focus, especially in reference to Black-crime stereotypes, several risk factors may simultaneously make discrimination more likely: for example, discretion, crime focus, and cognitive demand may have an additive effect in making errors based on stereotypicality more likely (e.g., Eberhardt et al., 2004; Kahn & Davies, 2011; Kahn et al., 2016). Specifically, the role of cognitive demand as a risk factor is that it prompts a reliance on cognitive shortcuts, such as stereotypes.

More precisely, a limited capacity for executive function means that people draw on the same resources for all operations that require this set of cognitive processes—including controlling one’s expression of bias, inhibiting other undesired actions, directing one’s attention and avoiding distraction, planning, choosing, reasoning, and holding things in working memory—each of which is more prone to error when the available resources are relatively scarce (e.g., Macrae, Bodenhausen, Milne, & Jetten, 1994; Miller, 2000; Robinson, Schmeichel, & Inzlicht, 2010). This can happen when cognitive load is relatively high, for example when a person is engaged in an especially difficult or taxing task while facing a simultaneous distraction that divides one’s attention. It can also occur when available cognitive resources are relatively low, such as when time pressure necessitates the same amount of information processing, despite less time to process the information. In these relatively demanding situations, people are more
likely to rely on simplifying assumptions such as stereotyping (e.g., Macrae, Milne, & Bodenhausen, 1994), which, again, heightens the risk of biased behavior.

**Time Pressure and Black-Crime Associations**

We have already seen that novice status and crime focus both amplify the effects of Black-crime associations, and next we will review how cognitively demanding situations can also act as risk factors. Black-crime associations are especially likely to result in discrimination in situations in which officers have to make decisions very quickly, a common situation encountered in modern patrol policing. Taking a closer look at relative cognitive resources as a situational risk factor for racially biased behavior, resource depletion is associated with increased reliance on stereotypes (Bodenhausen, 1990; Govorun & Payne, 2006; Macrae, Milne, et al., 1994; Pendry & Macrae, 1996). As just one example, the Implicit Association Test (IAT)—one relatively early standard for measuring implicit attitudes—was built on the premise of having limited time to respond, and on the premise of having to respond to more-versus-less cognitively demanding blocks of trials (e.g., trials with bias-inconsistent associative pairings, versus trials with bias-consistent pairings; Greenwald, Mcghee, & Schwartz, 1998).

As one example of time pressure as a risk factor for Black-crime associations resulting in discrimination, a program of research by Payne and colleagues, described in detail in the preceding section on crime focus, shows a bias toward perceiving weapons following Black face primes especially when participants are given a short amount of time to respond. When participants are given unlimited time to complete this task, they tend to correctly identify the guns and tools, regardless of whether the objects appear in the presence of a Black or White man’s face. Participants do typically identify guns faster when they accompany Black versus White men’s faces, suggesting that even with unlimited time, seeing Black men’s faces tends to
prime participants to perceive guns more quickly – however, time pressure intensifies these effects.

When given a short response deadline, participants are significantly more likely to mistake a tool for a gun when paired with a Black versus White man’s face. In other words, when faced with a split-second decision, race influenced participants’ tendency to perceive a gun that was not actually there (Payne, 2001, 2006). This evidence is reinforced by research that controls for image brightness and contrast using desaturated images that have been equalized for these properties, or by using line-drawing images (e.g., Eberhardt et al., 2004). While one set of explanations suggests people learn Black-crime associations from base rates of criminality in their social surroundings, given that people tend to hold these associations even as they pertain to Black children (Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014), base rates cannot fully explain this set of findings. Taken together, this research shows time pressure tends to increase the risk of anti-Black bias related to perceiving even non-existent weapons.

While these laboratory simulations are loose approximations of actual police work (often including pressing keyboard buttons on a computer rather than pulling a trigger to simulate shooting), some of these studies do suggest a stronger link to actual police behavior. A series of studies in field settings with police reveal patterns of racial disparities in behavior, including Eberhardt and colleagues’ (2004) study on suspect identification described above in the section on crime focus. In addition to demonstrating an effect of cognitively demanding situations, this particular example also shows one way in which crime focus and discretion simultaneously act as risk factors for stereotype-based errors.

**Divided Attention in Interracial Interactions**
In addition to time pressure, other situational factors can place demands on individuals’ cognitive capacity for controlling prejudiced behavior, including interracial interactions. If the stress of an interracial officer-civilian interaction were itself a risk factor for discrimination, that would be important to know – and research in laboratory settings indicates this is likely, though the research has not yet been conducted with officers. Recalling the effects of aversive racism from the section on discretion, interracial interactions can themselves sometimes result in adverse outcomes despite egalitarian intentions.

As shown in research by Shelton, Trawalter, and Richeson and their colleagues, one cause of depleted cognitive resources is the stress of interracial interactions (Richeson, Trawalter, & Shelton, 2005; Salvatore & Shelton, 2007; Shelton, Richeson, Salvatore, & Trawalter, 2005; Trawalter & Richeson, 2006). Interracial interactions can be stressful to non-White group members due to concerns about being the target of prejudice (Shelton et al., 2005) or about not being respected (Bergsieker, Shelton, & Richeson, 2010) while majority group members may be concerned about being perceived as racist, or about not being liked (Bergsieker et al., 2010)—all of which are experiences that tend to deplete available cognitive resources. In a paradigmatic example of these laboratory experiments involving actual face to face interactions, White participants perform worse on a subsequent measure of cognitive control after a conversation with a Black experimenter, compared to White participants who interacted with a White experimenter (Richeson et al., 2005). The greater their individual levels of implicit bias as measured by an IAT, the worse participants subsequently performed on the cognitive control task after an interracial (but not same race) interaction. Therefore, intergroup contact, thought to be a protective factor under conditions of equal status and shared goals (e.g., Pettigrew & Tropp,
206), can itself act as a risk factor in the short and medium term in its asymmetrical and fraught tendencies.

**Identity Salience and Identity Threats**

With public perceptions that racial disparities in policing reflect a broader problem of racial bias, it may not be surprising that a recent Pew study found fully 86% of officers believe that high-profile incidents of anti-Black violence have made their jobs harder (Morin et al., 2017). Similarly, 68% of officers in the same study believed protests following fatal shootings were motivated by anti-police bias. Whether or not officers’ jobs have become harder due to factors like decreased civilian compliance, officers’ subjective experience of being seen as prejudiced can also influence their cognition and behavior. What are officers’ day-to-day experiences of their social identities as police officers, and what are the consequences? The psychological literature on identity salience and threat does not yet include a robust literature on law enforcement officers, but some studies in these areas suggest ways that identity could matter in policing contexts.

**Identity Salience**

Research on identity salience is part of a larger literature on Social Identity and Self-Categorization Theories, in which individuals’ sense of identity ranges from one’s own personal identity, to their sense of belongingness in different social groups, known as collective identities (e.g., race, ethnicity, gender, religious group, nationality). These collective identities create the context for intergroup relations. When a group identity becomes salient, individuals tend to be more likely to focus on their group’s needs and goals, rather than their own individual needs and goals (Tajfel, Billig, Bundy, & Flament, 1971; Tajfel & Turner, 1979; Turner, Oakes, Haslam, & McGarty, 1994). While research has not examined the effects of different and overlapping
collective identities in police officers, identification with law enforcement is often an important aspect of an officer’s life (Bahn, 1984; Loftus, 2010; Rubin, 1972). This suggests that the psychological role of identity may be similar to other important identities studied in other laboratory and field settings.

The dynamic nature of social identities means that social situations can threaten the self concept by making a particular group membership salient. Occasionally, this dynamic property of social identities can be harnessed, resulting in less negative outcomes. For instance, Dovidio, Gaertner and colleagues have found that asking participants to think of themselves at broader levels of collective identity, thereby creating a “common ingroup identity,” can effectively shift the inclusiveness of the ingroup, and therefore reduce intergroup bias (Gaertner, Rust, Dovidio, Bachman, & Anastasio, 1994). As one example, students who were prompted to consider themselves as students in general, rather than as students of their particular, French-speaking university, were less likely to discriminate against students from a non-French speaking university (Yzerbyt, Dumont, Wigboldus, & Gordijn, 2003). In this and other experiments built on the common ingroup identity model, individuals’ motivation shifts toward the benefit of the currently salient, broader collective group. This idea may be one reason for the intuitive appeal of community-oriented policing programs: if officers identify more strongly with the broader community, an “us-versus-them” mentality may shift toward a collective “we” mentality, aligning officers’ and community members’ goals and outcomes.

**Identity Threats**

An identity threat refers to a social stimulus that causes an individual to feel concern that their value as, status of, or membership in an important social category may be reduced. For instance, an honors student being laughed at by a teacher may feel their identity as “smart” is
threatened in that situation. Here, we discuss two forms of identity threat that may be especially applicable to policing: stereotype threat and masculinity threat. Stereotype threat is a concern with being evaluated in terms of or confirming a negative stereotype about one’s group (Steele & Aronson, 1995). That concern often leads to negative performance in contexts where an individual faces a challenging task. Masculinity threat describes a tendency for some men to reassert their masculine self concept in the face of real or imagined challenges to one’s standing as positively masculine (Vandello, Bosson, Cohen, Burnaford, & Weaver, 2008).

**Stereotype threat.** The threat of confirming a negative stereotype about a group one belongs to, for example non-White and low-SES students’ concern about confirming stereotypes about these groups’ poor academic performance, actually tends to cause long-lasting detriments to performance in academics, athletics, the dynamics of interpersonal interactions, and a wide range of other identity-relevant domains, in both laboratory and field settings (see, e.g., Steele, 1997; Steele & Aronson, 1995).

In the first and most widely-replicated field experiment investigating stereotype threat, high-performing college students were given a set of difficult standardized test questions, and were instructed the test was either not diagnostic of their academic ability; or that the test was diagnostic of their academic ability. The “diagnostic” condition presented an identity threat to Black students: specifically, making salient the possibility that they might perform badly on this alleged test of their academic ability, thereby confirming negative stereotypes about Black students’ academic ability. While no group differences were observed in the control condition, in the threat condition, Black students solved significantly fewer problems correctly than White students—with these results adjusted for any differences in SAT scores.
Stereotype threat extends to many other domains—as just one example, in cross-race interactions, Whites’ concern about appearing racist causes more negative intergroup interactions (Goff et al., 2008), with another study showing the resulting discomfort tends to decrease interest in interracial interactions over time (Pearson et al., 2008). Studies of stereotype threat in real life settings observe detriments to performance in many different domains—anywhere a person’s identity can be threatened, for example, where an identity is underrepresented—including low-SES students’ academic performance (Croizet & Claire, 1998); women’s performance on math tests (Spencer, Steele, & Quinn, 1999), an effect which has been shown separately to be even stronger for women who strongly identify with their gender (Schmader, 2002); White men’s math performance when taking a test alongside members of a stereotypically higher-performing group (Aronson et al., 1999); White men’s athletic ability, and Black men’s “sports intelligence” (Stone, Lynch, Sjomeling, & Darley, 1999). Laboratory experiments also show strong evidence for how and why identity threats tend to hamper performance: the threat of confirming a negative group stereotype tends to raise blood pressure (Blascovich, Spencer, Quinn, & Steele, 2001) and tends to occupy the same short-term memory capacity that is needed to perform the academic tasks in question (Schmader & Johns, 2003; Schmader, Johns, & Forbes, 2008).

In the published literature, stereotype threat has not been tested in policing settings. However, in contexts where racial tensions between police and communities are high, it is reasonable to hypothesize concerns with appearing prejudiced may influence officer interpersonal performance. If this leads to more interpersonal disfluency, as it has in previous laboratory settings (Goff et al., 2008; Richeson & Shelton, 2003), this concern should motivate further research translating stereotype threat to the field of policing.
Masculinity threat. In the domain of gender, a person’s identity as a man is also subject to threat. Broadly, men are more likely to affirm their own masculinity after it has been threatened, in ways ranging from prejudice to discrimination and sometimes aggression. For example, a study of male college students showed that when threatened by receiving bogus personality test feedback indicating they themselves had feminine (vs. masculine) qualities, they exhibited heightened negative feelings toward effeminate-seeming (but not masculine-seeming) gay men on average (Glick, Gangl, Gibb, Klumpner, & Weinberg, 2007).

Threats to a person’s racial and gender identities are not mutually exclusive, as shown by two experiments in which experiencing racial discrimination tends to heighten the effects of masculinity threat for Black men, but not for White men (Goff, Di Leone, & Kahn, 2012). In these experiments, Black male college students affirmed male gender norms more strongly, and completed more pushups, thought to reflect an expression of dominance, compared to White male college students, after an episode of racial discrimination. The second study showed one reason White men may not have experienced the need to re-affirm their masculinity as strongly: they tended to compensate by reaffirming their relatively higher social status. From these experiments, we can see that identity threats can result in dominant physical behaviors, showing how such threats affect behavioral outcomes and performance, in addition to individual measures of prejudiced attitudes.

Identity Salience and Stereotype Threat. The salience of one’s different group identities can also shift the effects of stereotype threat, when one’s different group memberships are linked to different group stereotypes. Shih, Pittinsky, and Ambady (1999) conducted a laboratory study in which undergraduate Asian-American women completed a difficult math test. Based on cultural stereotypes in the United States of women’s low quantitative ability and Asian
Americans’ high quantitative ability, the researchers predicted that highlighting these students’ female identity would threaten their performance, while highlighting their Asian American identity would boost their performance. Indeed, participants whose female identity was salient performed significantly worse than participants in the control condition, and participants whose Asian American identity was salient performed significantly better than participants in the control condition. The researchers also ran this study in Vancouver, BC, where stereotypes of Asian Americans do not include superior quantitative ability, and found that Asian American identity salience in a culture lacking this stereotype did not predict a boost in performance. In other words, these effects are linked to stereotypes, and not the actual identities associated with those stereotypes.

**Future Directions and Potential Implications**

In this section, we suggest next directions in research that follow from open questions in the literature we have described. In almost all cases, we call for research on the phenomena identified in the psychological literature, applied to the specific kinds of situations and contexts in modern patrol policing we have identified as likely risk factors. Rigorous field research on these risk factors should be done before any attempt to intervene at scale, and yet field research on sensitive issues is challenging in contexts where practitioners may reasonably doubt researchers’ intentions or efficacy. A comprehensive guide on how to partner with police agencies on research is beyond the scope of the current article, but in broad terms, researchers and practitioners have many goals in common, and we consider it helpful to maintain a research focus on actionable findings that help police do an even better job of protecting public safety. In the short sections that follow, we present examples of research that would be especially helpful in policing contexts, within the categories of risk factors we have identified.
Future Research on Discretion and Bias in Policing

Little psychological research has directly examined the role of discretion on police behavior. Still, laboratory and some field studies on SDO, aversive racism, and stereotypicality provide a basis for believing that high levels of discretion in patrol officers leaves them vulnerable to disparate treatment of residents based on their race. While there may be trainings and policies that attenuate that risk, primary research on this topic is warranted.

Anecdotal evidence from the Austin Police Department (APD) is also consistent with this perspective. In a recent report by the Center for Policing Equity in collaboration with the Urban Institute and the White House Police Data Initiative, an analysis of APD’s vehicle stops revealed an unusually high “hit rate” following traffic stops (Goff, Lloyd, Geller, Raphael, & Glaser, 2016). A “hit rate” is the percentage of investigatory stops that returned illegal contraband, and is often used as a measure of police efficiency (Knowles, Persico, & Todd, 2001). While no comparative analyses were run, and no direct tests were conducted, the Chief of APD at the time attributed the relatively high rate (compared to other published hit rates) to having restricted the discretion of his officers to stop or inspect motor vehicles (Goff et al., 2016). Analyses of departments that feature these types of restrictions on officer behavior could provide natural experiments for scholars investigating the role of discretion across mechanisms for bias.

Future Research on Novice Status and Training in Officers

When has there been enough training, “practice,” or expertise developed to trust someone with the power to take away someone’s life or liberty—and to do so equitably? At a time when departments nationwide are implementing training programs with the explicit intention of reducing the effects of racial bias, research is needed to understand what forms of experience and training may improve outcomes. As we have seen, experience sometimes improves outcomes,
and sometimes may not. Though more experience may help officers make more accurate, less biased shooting decisions in laboratory simulations, longer time on the force also appears to increase levels of explicit prejudice. There are many possible explanations for these different kinds of effects, and more research is needed to understand the domains in which experience and training influence racial disparities in different ways. Factors such as relative staffing shortages may also influence the average experience levels of different departments, and these factors should also be explored.

Although it might appear otherwise from the rapid deployment of anti-bias training programs, little yet is known about the various possible effects of different forms of trainings. This set of research questions is especially time-sensitive due to recent calls by policymakers for anti-bias trainings, especially focusing on implicit bias. The authors of this article firmly support the rigorous design and testing of such trainings. In policy documents ranging from the President’s Task Force on 21st-Century Policing (Final Report of the President’s Task Force on 21st Century Policing, 2015) to the recently issued U.S. Department of Justice findings in Baltimore (Investigation of the Baltimore City Police Department, 2016), police departments are being urged to implement implicit bias training.

Several training protocols have emerged in response to demand, but these programs have not yet been assessed for their ability to produce a change in behavior—even by measuring behavior before and after training in a single location—much less by comparing the effects of randomly assigned different trainings. Given the political urgency and lack of scientific grounding, it is particularly important that research on training take place. Similarly, it is important to discern between trainings and mere information sessions because there is little evidence to suggest that a raising of consciousness about the existence of implicit bias will
produce significant changes in discriminatory behavior (Glaser, Goff, & Charbonneau, in press). Still, the distinction is rarely made in the emerging police training field, providing yet another opportunity for researchers to add professional clarity and advance the scientific literature.

Similarly, departments that require longer periods of community exposure before allowing officers to patrol on their own may provide opportunities for fruitful research. For instance, the New Haven Police Department requires officers to walk their assigned beat for a year before fully certifying them as officers (Buffa, 2013). Other departments mint officers after relatively brief stints in a training academy and even briefer field training assignments. Understanding the optimal length of these kinds of exposure protocols is an important question for both practitioners and scientists.

**Future Research on Crime Focus in Policing Contexts**

Research on training protocols would be most fruitfully conducted in tandem with research on crime focus and its associated racial stereotypes. While it is unlikely to discover a police department that does not encourage their officers to focus on crime reduction, there is wide variation in the degree to which officers are incentivized to have additional professional obligations. Specifically, the distribution of so-called “Community Policing” programs is particularly relevant. Though many bemoan that the term community policing is used to describe such diverse—and even contradictory programs (Greene & Mastrofski, 1988), department policies that promote engaging communities in positive and relationship-building contacts (as opposed to promoting crime deterrence) provide great opportunities to measure the influence of non-crime-focused intentions. For instance, requiring officers to meet a quota of positive contacts, rewarding officers who build community equity, and providing officers with opportunities to engage communities in recreational contexts may blunt some of the negative
consequences of a chronic crime-focused orientation. Each of these policies varies by departments (Cordner, 2014; Skogan & Hartnett, 1997), offering a natural opportunity to study their effects. A carefully designed study would both advance the science surrounding how mindsets influence stereotype application and help to address long-simmering debates about the influence of community policing on racial disparities in policing (Eck & Rosenbaum, 1994; Greene & Mastrofski, 1988).

**Future Research on Cognitive Demand in Policing Contexts**

Currently, though there is some work on time pressure and “shooting” decisions, there is no direct test of the role that divided attention, cognitive depletion, or interracial interactions have on officer-level racial biases. However, if particular policing policies or strategies increase the volume of police/citizen contacts, it is reasonable to wonder about the increased cognitive demand officers may incur as a result. In the other direction, it is also reasonable to wonder about whether training and expertise may attenuate the potentially biasing effects of time pressure and divided attention, and, again, what kinds of training may be more or less helpful. Further research is needed to translate the robust laboratory findings on cognitive demand, including time pressure and divided attention, into actual police patrol contexts.

**Future Research on Police Officer Identities**

Despite strong theoretical predictions and results from laboratory and field settings in a number of domains, to our knowledge there have not yet been studies of stereotype threat, masculinity threat, or identity salience as risk factors among police officers. Based on a long history of research on social identities and the motivational and cognitive consequences of identity salience and threats, we would suggest that future research address the possibility that different levels of identity might predict different consequences, especially in officer-civilian
interactions. In particular, threats to different aspects of an officer’s identity, for example threats to their social status, racial or gender identity, or to their sense of identity as a police officer, may put their performance at risk in the domain of keeping people safe. There is far too little research on the effects of police officers’ identities, and the resulting dynamics for interactions with civilians. For example, in interactions with people with different racial or gender identities, an officer’s own racial or gender identity may become more or less salient, influencing the course of the interaction. An officer’s collective identity with other police officers is likely relevant to other interpersonal outcomes. For example, when considering situations in which officers have more contact with people who may hold stereotypes of officers as being racist, and in which officers have increased discretion in managing that contact, the potential consequences of identity threats merit further investigation.

**Conclusion: A Call for Continued Research**

What can the psychological science of bias tell us about racial disparities in policing? At present, it is too early to know for certain. But in advance of achieving that certainty, the science does leave clues about where scientists can look to find evidence. In the preceding sections, we have attempted to map some of the most common experiences in contemporary policing to the robust risk factors for bias in the research literature in psychology. Again, these risk factors do not reflect on the character of individual officers. Rather, most reveal features of situations that make it more likely for any individual to engage in biased behaviors.

For the science of bias to speak compellingly to the realities of bias, more research must leave the laboratory and engage the world of policing. This research should be done carefully and in collaborative partnership with agencies that stand to benefit from its findings. Despite the fact that modern policing features a high degree of discretion, has to recruit novices, encourages
a chronic crime focus, exerts sizable cognitive demands, and promotes a strongly held identity, police departments vary widely in the cultures and policies that govern these experiences. As a result, policing is an ideal field laboratory for enterprising scientists convicted of the need to translate theory to experience.

While the present article may not present an exhaustive list of the experiences common to police patrols that may promote bias, these are among the experiences that most strongly align with robust risk factors in the psychological literature. A more exhaustive review would be welcome, but would also likely require more empirical evidence. Similarly, the present article does not articulate stages of police contact at which each experience might be most likely to influence officer behaviors (see Richardson & Goff, 2012, for a review).

What the present article has done is to demonstrate that, despite the lack of social psychological research on racial bias in policing, there is ample reason to suspect that social psychological factors play a role in policing outcomes. Moreover, there is great variability in the ways that departments regulate the experiences that endanger the fairness of officer behaviors. This natural variation is a topic worthy of study in and of itself, and a feature of the law enforcement landscape ripe for inquiry by social scientists. In the end, this is the strongest conclusion of the present review: That racial bias in policing is in dire need of increased scientific attention. Both practitioners and scholars have a great deal to gain from filling the gaps identified above—as do the uniformed officers sworn to keep us safe and the communities who most need to be able to rely on them.
Racial bias and policing

References


community-policing


Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and Explicit Prejudice and


Mekawi, Y., & Bresin, K. (2015). Is the evidence from racial bias shooting task studies a


http://doi.org/10.1111/1467-9280.00111


http://doi.org/10.1177/0146167212473157


