COMBATTING POLICE DISCRIMINATION IN THE AGE OF BIG DATA

Sharad Goel, Maya Perelman, Ravi Shroff and David Alan Sklansky*

[forthcoming in *New Criminal Law Review*]

[draft of May 31, 2016]

ABSTRACT

The exponential growth of available information about routine police activities offers new opportunities to improve the fairness and effectiveness of police practices. We illustrate the point by showing how a particular kind of calculation made possible by modern, large-scale datasets—determining the likelihood that stopping and frisking a particular pedestrian will result in the discovery of contraband or other evidence of criminal activity—could be used to reduce the racially disparate impact of pedestrian searches and to increase their effectiveness. For tools of this kind to achieve their full potential in improving policing, though, the legal system will need to adapt. One important change would be to understand police tactics such as investigatory stops of pedestrians or motorists as programs, not as isolated occurrences. Beyond that, the judiciary will need to grow more comfortable with statistical proof of discriminatory policing, and the police will need to be more receptive to the assistance that algorithms can provide in reducing bias.

Keywords: police, discrimination, stop-and-frisk, statistical proof, big data

* Sharad Goel is an Assistant Professor in the Department of Management Science and Engineering at Stanford University. Maya Perelman is a 2016 graduate of Stanford Law School and a law clerk for United States District Judge Jesus Bernal. Ravi Shroff is a Research Scientist at the New York University Center for Urban Science and Progress. David Alan Sklansky is the Stanley Morrison Professor of Law at Stanford Law School. The authors thank Sarah Brayne, John Donohue, Jeffrey Fagan, Michael Morse, Rob Park, James Rothschild, Andrew Selbst, and Robert Weisberg for helpful comments and criticism. This work was partially supported by the John S. and James L. Knight Foundation.

INTRODUCTION

The "second information revolution" has yet to transform law enforcement as dramatically as home telephones and radio dispatch systems did,[1] but that may be just a matter of time. Police departments have access to exponentially increasing amounts of information, and methods of processing and analyzing vast sets of data grow ever more sophisticated with dizzying speed. Big data is coming to policing.

A small but growing number of scholars have begun to address the important questions and profound difficulties that data mining by law enforcement agencies poses for traditional doctrines of criminal procedure.[2] We wish in this Article to highlight an equally important set of issues that big data raises for policing: not the challenge of constructing legal protections against the new powers that big data will give the police, but the new strategies it makes possible for overseeing the police and improving the fairness of law enforcement. Big data in policing is like the introduction of body cameras, writ large: it creates new threats to liberty and privacy, but it also offers new ways to hold the police accountable. The threats posed by police use of big data are, appropriately, receiving an increasing amount of scholarly attention. We want to draw

[1] GERALD W. BROCK, THE SECOND INFORMATION REVOLUTION (2004); ROBERT M. FOGELSON, BIG CITY POLICE (1977).

[2] *See, e.g.*, BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING AND PUNISHMENT IN AN ACTUARIAL AGE (2007); Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion,* 163 U. PA. L. REV. 327 (2015); Elizabeth E. Joh, *Policing By Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35 (2014). For discussion of related issues in the context of employment discrimination, see Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. ___ (2016).

similar attention to the opportunities that big data provides for strengthening police accountability and improving police practices.[3]

We focus here on a particularly important dimension of police accountability—guarding against racial discrimination—in a particularly important area of police activity—the "stop-and-frisk." The stop-and-frisk is also known as the "investigatory stop," the "pedestrian stop," or the *"Terry* stop." The last term comes from *Terry v. Ohio*,[4] the case in which the Supreme Court first addressed, nearly half a century ago, the compatibility of this police tactic with the Fourth Amendment's ban on "unreasonable searches and seizures." *Terry* stops have always been controversial, perhaps first and foremost because of concerns that they are employed excessively against members of racial minorities. The *Terry* case was decided three months after the National Advisory Commission on Civil Disorders identified "field interrogations and the 'stop-and-frisk' technique" as one of the leading contributors to the wide-scale urban riots of 1967,[5] and the Supreme Court itself noted the connection between the stop-and-frisk and "the wholesale

[3] For earlier explorations of some of the themes we address here, see Mary D. Fan, *Panopticism for Police: Structural Reform Bargaining and Police Regulation by Data-Driven Surveillance*, 87 WASH. L. REV. 93 (2012), and Steven Rushin, *Using Data to Reduce Police Violence*, 57 B.C. L. REV. 117 (2016). Fan focuses on the potential for big data to support oversight of the police by "politically attuned agencies and civil societies," Fan, *supra*, at 97; we are more concerned with how big data can reinvigorate judicial remedies for police discrimination. Nonetheless the new statistical techniques we describe can strengthen the kinds of non-judicial uses Fan discusses, as well. Fan shares our hope that big data will prod police departments toward "self-examination and change" on their own accord. *Id.* at 129. Rushin discusses how the United States Department of Justice could use data regarding police shootings in determining which police agencies it should consider suing for civil rights violations, to shame departments into reforming, and to monitor the effectiveness of reform efforts; he is skeptical about the ability of "data transparency" to spur reform "without additional top-down" incentives. Rushin, *supra*, at 136.

[4] 391 U.S. 1 (1969).

[5] REPORT OF THE NATIONAL ADVISORY COMMITTEE ON CIVIL DISORDERS 143-44, 302 (1968); *see* David A. Sklansky, *Traffic Stops, Minority Motorists, and the Future of the Fourth Amendment*, 1997 SUP. CT. REV. 271, 314-15.

harassment by certain elements of the police community, of which minority groups, particularly Negroes, frequently complain."[6] Nonetheless, the Court upheld the constitutionality of the stop-and-frisk, as long as it was supported by "specific and articulable facts" justifying suspicion of criminality and danger[7]—a standard the Court later restated as "reasonable, articulable suspicion,"[8] or simply "reasonable suspicion."[9]

One advantage of requiring that the basis of suspicion be "articulable," as opposed to an "inchoate and unparticularized . . . 'hunch,'"[10] is that it can help to smoke out "illegitimate motive[s]."[11] In particular, if officers have to explain the basis for a stop, there should be less room for stops motivated by racial prejudice. Nonetheless racial bias has continued to plague *Terry* stops. In 2013, federal plaintiffs challenging the stop-and-frisk practices of the New York Police Department in a case called *Floyd v. City of New York* obtained a broad structural injunction based on findings by the trial judge that blacks and Hispanics were stopped disproportionately and unjustifiably.[12] Two years later the United States Department of Justice released its report on policing in Ferguson, Missouri. That report blasted Ferguson for a pattern

---

[6] *Terry*, 391 U.S. at 10.
[7] *Id.* at 21.
[8] Brown v. Texas, 443 U.S. 47, 51 (1979); *accord, e.g.*, Illinois v. Wardlow, 528 U.S. 119, 123 (2000).
[9] *E.g.,* Rodriguez v United States, 575 U.S. ___, 135 S. Ct. 1609, 1613 (2015); United States v. Brignoni-Ponce, 422 U.S. 873, 882 (1975).
[10] *Terry*, 391 U.S. at 27; *see also id.* at 22.
[11] *See, e.g.*, United States v. Fong, 662 F. Supp. 1319, 1322 (D. Del. 1987)
[12] Floyd v. City of New York, 959 F. Supp. 2d 540 (S.D.N.Y. 2013). We use the term "Hispanic" throughout this Article because that is the term used in the NYPD's records. For similar reasons, we use "black" instead of "African American." We note, though, that many people of Latin American descent prefer the term "Latino," and some object to either term. *See, e.g.*, PAUL TAYLOR ET AL., PEW RESEARCH CENTER, WHEN LABELS DON'T FIT: HISPANICS AND THEIR VIEWS OF IDENTITY (2012), *available at* <www.pewhispanic.org>; Cindy Y. Rodriguez, *Which Is It, Hispanic or Latino?*, CNN, May 3, 2014, *available at* <www.cnn.com>.

of racially discriminatory policing, and suggested that investigatory stops of pedestrians in Ferguson probably reflected the broader pattern, although that could not be determined with certainty because Ferguson did "not track or analyze" those stops "in any reliable way."[13] Studies of a related police practice, also governed by *Terry v. Ohio*—investigatory stops of automobiles for purposes of discovering criminal activity—have repeatedly found evidence of racial bias.[14]

The Supreme Court noted in *Terry* that the stop-and-frisk is "a serious intrusion upon the sanctity of the person, which may inflict great indignity and arouse strong resentment."[15] Indeed, the very term "stop-and-frisk" struck the Court as a "euphemis[m]."[16] We are talking, the Court stressed, about a police officer accosting an individual, restraining his liberty to leave, and engaging in "a careful exploration of the outer surfaces of a person's clothing all over his or her body . . . while the citizen stands helpless, perhaps facing a wall with his hands raised."[17]

Even that description falls well short of capturing the violence and humiliation often associated with *Terry* stops in minority neighborhoods.[18] There is good reason to believe that the stop-and-frisk tactics*, especially* when they are thought to be racially targeted, take a heavy toll on the perceived legitimacy of the legal system in minority communities; that in turn undermines

---

[13] U.S. DEP'T OF JUSTICE, CIVIL RIGHTS DIV., INVESTIGATION OF THE FERGUSON POLICE DEPARTMENT (2015).
[14] *See, e.g.* CHARLES R. EPP, STEVEN MAYNARD-MOODY & DONALD HAIDER-MARKEL, PULLED OVER: HOW POLICE STOPS DEFINE RACE AND CITIZENSHIP 3 & 185 n.10 (2014).
[15] *Terry*, 391 U.S. at 17.
[16] *Id.* at 10.
[17] *Id.* at 16-17.
[18] *See, e.g.*, VICTOR M. RIOS, PUNISHED: POLICING THE LIVES OF BLACK AND LATINO BOYS 4-5, 125-26, 147, 149, 155 (2011).

voluntary compliance with the law as well as the trust and cooperation required for effective policing.[19]

Racial discrimination in *Terry* stops has persisted in part because it is often difficult to detect and in part because legal remedies for it are weak. The two problems are related: part of the weakness of the legal remedies lies in the kinds of proof that they require. True, a stop-and-frisk is illegal under *Terry* if the officer cannot identify specific facts justifying the intrusion, but a minimally inventive officer can almost always find a facially legitimate reason to suspect someone the officer wishes to frisk. In a pinch, there is always something like, "he looked nervous and acted evasive." Furthermore, the Supreme Court has made clear that as long as objective facts justify an investigatory stop, the officer's actual motivation is irrelevant under the Fourth Amendment.[20] Motive *does* matter under the Equal Protection Clause of the Fourteenth Amendment, but establishing an equal protection violation requires proof of "discriminatory intent," a legal term of art that means, more or less, animus directed at a particular racial minority or other protected class.[21] Few officers admit to animus of that kind.

One way around these problems of proof is to rely on statistical analysis of large numbers of stops, but efforts along these lines have encountered two obstacles, one practical and the other jurisprudential. The practical problem has been that detailed information about *Terry* stops is

---

[19] *See, e.g.,* Aziz Z. Huq, Tom R. Tyler & Stephen J. Schulhofer, *Why Does the Public Cooperate With Law Enforcement?*, 17 PSYCHOL. PUB. POL'Y & L. 419 (2011); Benjamin Justice & Tracey L. Meares, *How the Criminal Justice System Educates Citizens*, 651 ANNALS AM. ACAD. POL. & SOC. SCI. 159 (2014); Tom R. Tyler & Jonathan Jackson, *Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement*, 20 PSYCHOL. PUB. POL'Y & L. 7 (2014).

[20] *See* Whren v. United States, 517 U.S. 806 (1996).

[21] *See, e.g.*, Washington v. Davis, 426 U.S. 229 (1976).

often limited. Investigatory stops are often rapid and relatively informal; they do not generate the kind of extensive records produced for a nonconsensual search of a house, or for a wiretap. The Ferguson Police Department is hardly unique in having failed to "track or analyze pedestrian *Terry* stops . . . in any reliable way." The jurisprudential problem is that, even when statistical evidence *is* available, courts often treat it with hostility and suspicion. The Supreme Court itself has set the tone. Three decades ago, in *McCleskey v. Kemp*,[22] the Court upheld Georgia's death penalty in the face of striking evidence of racial bias: defendants convicted of killing white victims were four times more likely to receive a death sentence than defendants convicted of killing blacks, and blacks who killed whites were sentenced to death seven times more frequently than whites who killed blacks. The Court treated these statistics as largely irrelevant, because they could not show any *particular* defendant had been sentenced to death because of his race or the race of his victim. Lower courts have taken their lead from *McCleskey* in cases challenging racial discrimination virtually anywhere in the criminal justice system. What matters, the courts say, is what happened in a particular case, not the behavior of the system overall.

Big data, together with the algorithmic tools that help us draw insights from such data, is making inroads into the first of these obstacles, and it may make the second less prohibitive. Technology has made it easier and easier to record, collect, and analyze data on *Terry* stops, and more and more police departments are doing so—sometimes under judicial prodding, and sometimes voluntarily.[23] Increasingly, too, this data is being made available to researchers outside of law enforcement, enabling them to conduct increasingly sophisticated statistical tests

[22] 481 U.S. 279 (1987).
[23] David A. Harris, *Across the Hudson: Taking the Stop and Frisk Debate Beyond New York City*, 16 N.Y.U. J. LEGIS. & PUB. POL'Y 853, 856, 869-71 (2014).

for racial bias.[24] The New York Police Department has long taken the lead in compiling stop-and-frisk data, and data it collected provided much of the basis for the path-breaking remedial injunction entered in *Floyd*. That decision, we will argue, is a sign of things to come. As statistical proof of racial bias becomes more and more convincing, courts will—and should—be more comfortable relying on it. So will, and should, forward-thinking police departments, acting on their own initiative.

The *Floyd* case, in fact, falls short of demonstrating the full power that big data offers, even today, to detect and reduce racial discrimination in stop-and-frisk practices. Judge Shira Scheindlin's opinion in *Floyd* was based in part on statistical evidence showing that blacks and Hispanics were more likely than whites to be subjected to *Terry* stops in New York City, and that many of the stops were justified, according to the NYPD's own records, by factors that correlated negatively with the likelihood that the stop would turn into an arrest. Those statistics alone, though, were insufficient to show that any particular stop was unconstitutional. And to find that the NYPD's *Terry* policy as a whole violated equal protection, Judge Scheindlin found

[24] *See, e.g.*, Joscha Legewie, *Racial Profiling in Stop-and-Frisk Operations: How Local Events Trigger Periods of Increased Discrimination*. AMERICAN JOURNAL OF SOCIOLOGY, forthcoming (2016) (comparing similar police stops before and after acts of violence towards police officers and finding an increase in use of physical force against blacks after fatal shootings of NYPD officers by black suspects, but no corresponding increase in force after fatal shootings of NYPD officers by white or Hispanic suspects); Greg Ridgeway & John MacDonald, *Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops*. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 104 (2009) (finding that of the nearly 3,000 New York City Police Department officers regularly involved in pedestrian stops, 15 officers stopped a substantially greater fraction of black and Hispanic suspects than their statistical benchmark predicts); Andrew Gelman, Jeffrey Fagan and Alex Kiss, *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 102 (2007) (finding that minorities were stopped more often than whites, both in comparison to their proportion in the local population, and relative to local crime rates in those groups).

it necessary to rely on evidence—not usually available in a case of this type—that the police department had explicitly instructed its officers to target young black and Hispanic men for stop-and-frisks.

But the data on *Terry* stops in New York City can be used to show more than what the *Floyd* opinion indicated. The data can be used to compute the likelihood that any particular stop-and-frisk will result, for example, in the discovery of particular kinds of evidence, given the information available to the officer before the encounter—i.e., time of day, location, suspect characteristics, and the circumstances identified by the officer as giving rise to suspicion.[25] All of this information is recorded in what the NYPD calls a "UF-250" report, and it can be used to estimate a "stop-level hit rate"[26]—the *ex ante* probability of discovering a weapon, based on all the factors that were known to the officer before the *Terry* stop. The stop-level hit rate, or "SHR," can be thought of as a measure of the strength of the evidence supporting the suspicion that the individual to be stopped and frisked has a gun. According to the NYPD's records, close to half a million *Terry* stops in New York City between 2008 and 2010 were based on suspicion of "criminal possession of a weapon," or "CPW"; for these stops, the SHR provides a kind of numerical measure of "articulable suspicion."[27]

SHRs computed in this manner are quite revealing. It turns out that 43% of the *Terry* stops carried out by the NYPD based on suspicion of CPW had less than a 1% chance of actually resulting in the discovery of a weapon. And these low-odds stops had a heavy racial tilt: 49% of

[25] Sharad Goel, Justin M. Rao & Ravi Shroff, *Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy*, ___ ANN. APP. STATS. ___ (forthcoming).
[26] *Id.* at ___.
[27] *See id.* at ___.

the stops of blacks fell below the 1% probability threshold, and 34% of the stops of Hispanics, but only 19% of the stops of whites.[28] So the SHR method offers further support for the finding in *Floyd* that the stop-and-frisk practices of the NYPD have been racially discriminatory.

But the SHR method does more than that. It pinpoints particular categories of *Terry* stops for CPW that both (a) are relatively unlikely to actually find a weapon, and (b) impose an especially disproportionate burden on racial minorities. The SHRs can thus provide a road map for redirecting stop-and-frisk practices to make them, simultaneously, less racially lopsided in their impact and more effective at finding what the police say they are looking for. SHR analysis reveals that some of the standard justifications for pedestrian stops that the UF-250 has employed—"furtive movements," for example—are unhelpful in identifying suspects who actually have weapons[29]; avoiding the use of those factors would make stops less discriminatory and more successful. More ambitiously, SHR analysis could be used to craft a simple heuristic for officers to use on the street to determine which suspects to stop and frisk, drastically reducing the disparate impact and increasing the "efficiency" of the searches.[30] Bernard Harcourt and Tracey Meares have argued provocatively that *Terry* stops should be randomized to avoid invidious discrimination.[31] The SHR method could provide equivalent protection against discrimination while preserving the use of individualized suspicion—and improving the effectiveness of the stops in carrying out their stated objective.

---

[28] *See id.* at ___.
[29] *See id.* at ___. Appropriately, the current version of the UF-250 removes "furtive movements" as a justification.
[30] *See* Sharad Goel, Justin M. Rao & Ravi Shroff, *Personalized Risk Assessments in the Criminal Justice System*, ___ AM. ECON. R. ___ (forthcoming).
[31] Bernard E. Harcourt & Tracey L. Meares, *Randomization and the Fourth Amendment*, 78 U. CHI. L. REV. 809 (2011).

Because the SHR method provides an objective, numerical measure of the *ex ante* likelihood of discovering what the police say they are looking for, it also could be used to help assess the existence *vel non* of "reasonable articulable suspicion" in particular cases. The Supreme Court has been famously insistent that the legal standards for searches and seizures cannot be quantified or reduced to mechanical formulas,[32] but the SHR method has a critical advantage in this regard: it relies on the factors that the police department itself has identified as relevant in justifying a *Terry* stop. So if the police believe that the SHRs calculated from their stop data do not fully capture the strength of suspicion in some cases, they can—and will have an incentive to—revise the form officers use to record the basis for a stop-and-frisk. Testing for reasonable suspicion with SHRs calculated from the police department's own form is thus a way of using big data to give bite to the requirement that *Terry* stops be based on "specific and articulable facts" rather than mere hunches.

For statistical tools like the SHR method to achieve their full potential in reducing police discrimination, legal doctrine will need to adapt. The most important change would be to understand *Terry* stops—both the pedestrian stops on which we focus here, and the related tactic of investigatory traffic stops—as programs, not as isolated occurrences.[33] Stop-and-frisk *is* a program, and it always has been. It was conceived as an organized tactic by police departments

[32] *See, e.g.*, Maryland v. Pringle, 540 U.S. 366, 371 (2003); Illinois v. Gates, 462 U.S. 213 (1983).
[33] This is a point that Tracey Meares has rightly stressed. *See* Tracey L. Meares, *Programming Errors: Understanding the Constitutionality of Stop-and-Frisk as a Program, Not an Incident*, 82 U. CHI. L. REV. 159 (2015).

in the 1960s,[34] and it is carried out today, generally, by officers acting pursuant to departmental policies.[35] The same is true of investigatory traffic stops: they are "scripted, predictable, and deeply institutionalized … not the creation of individual officers' discretionary decisions but of organized training, professional norms, and shared expectations."[36] The harms inflicted by *Terry* stops, moreover, are often tied to their programmatic nature: the intrusions give rise to resentment in minority communities in large part because they are experienced as cumulative, and because they are understood to be racially targeted.[37] Furthermore, big data is itself heightening the programmatic nature of police stops, as law enforcement agencies increasingly turn to databases and number-crunching to target their enforcement strategies.[38]

The Supreme Court nonetheless analyzed the police actions challenged in *Terry v. Ohio* as if they were decisions by a lone police officer acting on his own initiative, and that conceptual frame has stuck.[39] At the time *Terry* was decided, there was perhaps an argument for that frame: treating stop-and-frisk as a program, and assessing its reasonableness as a program, would have

---

[34] *See, e.g.*, CHRISTOPHER LOWEN AGEE, THE STREETS OF SAN FRANCISCO: POLICING AND THE CREATION OF A COSMOPOLITAN LIBERAL POLITICS, 1950-1972, at 35-39 (2014); FOGELSON, *supra* note 1, at 187-88.
[35] *See* Meares, *supra* note 33.
[36] EPP, MAYNARD-MOODY & HAIDER-MARKEL, *supra* note 14, at 36.
[37] *See, e.g., id.* at 6.
[38] *See, e.g.*, Ferguson, *supra* note 2; Maurice Chammah, *Policing the Future*, THE MARSHALL PROJECT, Feb. 3, 2016, *available at* <themarshallproject.org>.
[39] Nine years after deciding *Terry*, the Court suggested in dicta that a stop-and-frisk in the absence of reasonable suspicion might be lawful if carried out "pursuant to a practice embodying neutral criteria." Brown v. Texas, 443 U.S. 47, 51 (1979). But unless a state chose to defend its stops on that ground—and, overwhelmingly, states have not made that choice—the Court has analyzed *Terry* stops as isolated occurrences. In this respect stop-and-frisk doctrine is typical of Fourth Amendment law more generally, which Daphna Renan points out has long been "transactional" in orientation, focusing on "discrete law enforcement-citizen encounter[s]." Daphna Renan, *The Fourth Amendment as Administrative Governance*, 68 Stan. L. Rev. ___, ___ (2016).

required delving into the internal, departmental decision-making of the police. That kind of inquiry might have struck the Court as too intrusive. Today, though, big data is both heightening the programmatic nature of *Terry* stops and allowing increasingly sophisticated assessments of those programs based on objective measures of how they are actually carried out.[40] Even the SHR method, which allows for tailored measurements of the grounds for suspicion in individual cases, is most powerful as a tool for assessing—and improving—departmental policies. Given the increasing power of tools such as SHR, it makes less and less sense to assess the constitutionality of a *Terry* stop by focusing solely on its particular circumstances, without regard to the policy decisions shaping the department's overall stop-and-frisk program. In particular, we will argue, the reasonableness of a stop-and-frisk under the Fourth Amendment, as well as its compatibility with the Equal Protection Clause of the Fourteenth Amendment, should depend in part on how police departments respond, or fail to respond, to what big data demonstrates about the department's policies and practices.

Part I of this Article will describe the difficulties that litigants traditionally have encountered in seeking to use either the Fourth Amendment requirement of "reasonable articulable suspicion" or the Fourteenth Amendment command of equal protection to challenge racial discrimination in *Terry* stops. Part II will use the NYPD and the *Floyd* case to explore how big data is beginning to provide new tools for combatting police discrimination. Part III will discuss the SHR method and its potential for reducing police discrimination while improving the effectiveness of *Terry* stops. Part IV will discuss the lingering obstacles that existing doctrines create for making full use of big data in general and SHRs in particular to combat police

[40] *See* Meares, *supra* note 33.

discrimination, and the importance of understanding stop-and-frisk as a program rather than a series of isolated encounters.

I. Police Discrimination and the Constitution

Constitutional law provides two ways of challenging racial discrimination in a police department's use of stop-and-frisk powers. The first is to use the Fourth Amendment standard of "reasonable articulable suspicion" to smoke out illegitimate motive, and the second is to raise an equal protection claim under the Fourteenth Amendment. In order to appreciate the importance of big data in improving the racial fairness of *Terry* stops, it will help to begin with an exploration of the difficulties that litigants—whether individual criminal defendants, private civil plaintiffs, or the federal government seeking a structural injunction[41]—have traditionally experienced in seeking to use either of these doctrinal tools to challenge police discrimination.

A. Reasonable Articulable Suspicion

Prior to the Supreme Court's 1968 decision in *Terry v. Ohio,* it was unclear whether the Fourth Amendment applied to seizures of persons short of arrest, and if so, what requirements it

[41] 42 U.S.C. § 14141, enacted in 1994, authorizes the Department of Justice to sue in federal court for a structural injunction against police departments engaged in a "pattern or practice" of constitutional violations; it provides an additional and very important *remedy* for police discrimination, but does not change what is required in order to *prove* that police discrimination amounts to a constitutional violation. *See generally* Rachel Harmon, *Promoting Civil Rights Through Proactive Police Reform*, 62 STAN. L. REV. 1 (2009); Steven Rushin, *Federal Enforcement of Police Reform*, 82 FORDHAM L. REV. 3189 (2014). Much of the work done by § 14141 has come in the form not of injunctions achieved after court battles but of consent decrees and out-of-court settlements reached between police departments and the Department of Justice. *See* Fan, *supra* note 3, at 116-24.

imposed. That doctrinal silence became increasingly intolerable as police departments in the 1960s began making systematic use of the stop-and-frisk technique as a tool of order maintenance and crime control, and minority communities began complaining that the technique was a racially targeted instrument of harassment. When the Court finally addressed the stop-and-frisk in *Terry*, it neither declared the tactic outside the scope of the Fourth Amendment (as law enforcement had requested) nor declared it flatly unconstitutional (as many civil libertarians urged). Instead, the Court steered a middle course, allowing police officers to temporarily detain a person on the basis of reasonable suspicion that "criminal activity may be afoot," and to conduct a "carefully limited search" of the outer clothing to check for weapons, if there was reason to think the person was armed.[42] To find that the intrusion was reasonable, the Court employed the traditional balancing test required by the Fourth Amendment, which weighs the need to search or seize against the invasion which the search or seizure entails.[43] Ultimately, despite its recognition that such a stop is "a serious intrusion upon the sanctity of the person," the Court found that such stops were part of police officers' proper investigative function.[44] Furthermore, in light of the necessity for "swift action predicated upon the on-the-spot observations of the officer on the beat," the Court dispensed with the requirement of probable cause as imposing an impractical limitation on effective crime prevention and detection.[45]

---

[42] 391 U.S. at 16-17. Writing for majority in *Terry*, Chief Justice Warren thought it was unclear whether the defendants had been detained before being patted down, so he expressly declined to address the legal requirements for a "stop" without a "frisk." *See id.* at 19 n.16. Justice Harlan's concurrence, though, argued that a "stop" requires reasonable suspicion of criminality, while a "frisk" requires both a lawful stop and reasonable suspicion that the suspect is armed and dangerous. *See id.* at 32-33. That is the governing law today. *See, e.g.*, Minnesota v. Dickerson, 508 U.S. 366, 372-73 (1993). Nonetheless, the terms "*Terry* stop" and "pedestrian stop" are often used to describe a stop combined with a frisk.

[43] *Id*. at 21.

[44] *Id*. at 17, 22.

[45] *Id*. at 20-22.

However, Chief Justice Warren's opinion made clear that the lower standard did not negate the requirement of specified and articulable suspicion, emphasizing, "This demand for specificity in the information upon which police action is predicated is the central teaching of this Court's Fourth Amendment jurisprudence."[46] While a *Terry* stop dispenses with the role of the neutral magistrate in approving a search or seizure, he emphasized that the principle behind warrant applications still applies; for a stop to comply with the Fourth Amendment, an officer must be able to articulate the factual basis for his suspicion. In the case of defendant Terry himself, the Court found that the officer had met his burden where the defendant and his companion paced back and forth past a store window, followed by a conference with a third man; under these facts, the Court agreed with the officer's assessment that the men's behavior was sufficiently suggestive of a preface to a stick-up.[47]

Since *Terry*, the Court has repeatedly reaffirmed the requirement of individualized suspicion.[48] The Court has dispensed with that requirement only in what it used to call the "closely guarded category"[49] of "special needs" cases.[50] That category is a good deal more

---

[46] *Id*. at 20, n. 18.
[47] *Id*. at 16-17.
[48] *See, e.g*., Almeida-Sanchez v. United States, 413 U.S. 266 (1973) (stop made without reasonable suspicion as part of roving police patrol unreasonable under Fourth Amendment, and thus unconstitutional); United States v. Brignoni-Ponce, 422 U.S. 873 (1975) (border patrol not entitled to dispense entirely with the requirement that officers must have a reasonable suspicion to justify roving patrol stops); Delaware v. Prouse, 440 U.S. 648, 663 (1979) (random stop to check driver's license and registration unreasonable in absence of articulable and reasonable suspicion that motorist is unlicensed or that automobile not registered); City of Indianapolis v. Edmond, 531 U.S. 32, 34 (2000) (highway checkpoint program whose primary purpose was discovery and interdiction of illegal narcotics violated Fourth Amendment because general interest in crime control does not justify regime of suspicionless stops).
[49] Ferguson v. City of Charleston, 532 U.S. 67, 74 (2001).

expansive than it once was, but it remains limited to searches—like sobriety checkpoints, drug tests in public schools, and searches of probationers—that address concerns beyond the policing of "ordinary criminal wrongdoing."[51] In the context of stop-and-frisk, where the rationale for stops is based squarely on crime detection and prevention, the law thus remains that officers must have particularized suspicion to justify a stop.

Litigants have found it difficult, though, to challenge *Terry* stops on the ground that particularized suspicion was lacking. As articulated by the Court, the reasonable suspicion standard "falls considerably short of satisfying a preponderance of the evidence standard"[52] and entails only "some minimal level" of justification.[53] Furthermore, a determination that reasonable suspicion exists "need not rule out the possibility of innocent conduct."[54] Rather, courts have established that the test involves reviewing the totality of the circumstances at the time of the stop.[55] As well, in applying this test, the Court has made it clear that the standard is a lenient one, satisfied by an officer's recital of any of a (facially neutral) laundry list of factors. In *United States v. Brignoni-Ponce*, the Court set out an extensive—but not exclusive—list of such factors, which included the characteristics of the area, the type of vehicle being used, the number of

---

[50] *Id.* at 68 (2001) (acknowledging case law carving out "special needs" exceptions to individualized suspicion requirement where government raises concerns outside of normal law enforcement").
[51] *Edmond*, 531 U.S. at 34.
[52] United States v. Arvizu, 534 U.S. 266, 274 (2002)
[53] United States v. Sokolow, 490 U.S. 1, 2 (1989)
[54] 534 U.S. at 277.
[55] United States v. Cortez, 449 U.S. 411, 417-18 (1981) ("Courts have used a variety of terms to capture the elusive concept of what cause is sufficient to authorize police to stop a person. Terms like "articulable reasons" and "founded suspicion" are not self-defining; they fall short of providing clear guidance dispositive of the myriad factual situations that arise. But the essence of all that has been written is that the totality of the circumstances—the whole picture—must be taken into account.")

passengers in the car, and the manner of the suspect's dress or haircut.[56] In subsequent cases, the Court found that an individual's "evasive behavior" in retreating from the police[57] and his conformity to government-created criminal profiles[58] were similarly probative.

[56] 422 U.S. at 884-85. The full list is as follows: "Officers may consider the characteristics of the area in which they encounter a vehicle. Its proximity to the border, the usual patterns of traffic on the particular road, and previous experience with alien traffic are all relevant . . . They also may consider information about recent illegal border crossings in the area. The driver's behavior may be relevant, as erratic driving or obvious attempts to evade officers can support a reasonable suspicion . . . Aspects of the vehicle itself may justify suspicion. For instance, officers say that certain station wagons, with large compartments for fold-down seats or spare tires, are frequently used for transporting concealed aliens . . . The vehicle may appear to be heavily loaded, it may have an extraordinary number of passengers, or the officers may observe persons trying to hide . . . The Government also points out that trained officers can recognize the characteristic appearance of persons who live in Mexico, relying on such factors as the mode of dress and haircut . . . In all situations the officer is entitled to assess the facts in light of his experience in detecting illegal entry and smuggling." *Id*. The Court further agreed with the government that Mexican appearance was a relevant factor in determining the reasonableness of a stop, at least in determining whether he was illegally in the country, although it held that this factor could not, standing alone, be sufficient. *Id*. at 886-87. The Ninth Circuit, though, has determined that the likelihood that any given person of Hispanic ancestry is in fact an illegal alien is not high enough to make Hispanic appearance a relevant factor in the reasonable suspicion calculus. United States v. Montero-Camargo, 208 F.3d 1122, 1132 (9th Cir. 2000). In reaching this decision, the court reasoned that *Brignoni-Ponce* had relied on now-dated demographic information, and the "Hispanic appearance" characteristic "cast too wide a net" to play a part in the particularized suspicion decision in light of the increase in the Hispanic population in California since the Supreme Court's decision. *Id*. at 1133.

[57] Illinois v. Wardlow, 528 U.S. 119, 124 (2000) ("[W]e have previously noted the fact that the stop occurred in a "high crime area" among the relevant contextual considerations in a *Terry* analysis.")

[58] United States v. Sokolow, 490 U.S. 1, 2, 10 (1989) (affirming constitutionality of stop where defendant matched characteristics of Drug Enforcement Agency's drug courier profile). As Justice Marshall noted in his dissent, officers' ability to claim reliance on a profile broadened the range of behavior that could give rise to reasonable suspicion, a risk enhanced by the profile's "chameleon-like way of adapting to any particular set of observations." *Id*. at 13-14 (internal quotation marks omitted). As examples, he noted that officers had in the past relied on profiles to claim that the following all contributed to a reasonable suspicion determination: being the first to deplane, being the last to deplane, buying one-way tickets, buying round-trip tickets, taking a nonstop flight, changing planes, flying with a gym bag, flying with new suitcases, traveling alone, traveling with a companion, acting nervously, and acting too calmly. *Id*.

The Court's typical deference towards police officers' judgment further tips the scales in favor of finding for the constitutionality of a stop. Dating back to *Terry*, the Court's opinions have accorded great weight to police officer expertise in crediting their determinations of suspicious activity.[59] In *United States v. Cortez*, the Court made its confidence in officer judgment explicit, noting that police draw inferences from various observations "that may well elude an untrained person."[60] The Court further emphasized that the process of forming suspicion "does not deal with certainties, but with probabilities," and noted that officers' observations must be weighed "not in terms of library analysis by scholars, but as understood by those versed in the field of law enforcement."[61] Lower courts have followed the Court's lead in deferring to officer judgment. In denying a criminal defendant's motion to suppress, for example, an Ohio court of appeals repeated the language from *Cortez*, noting that "due deference must be accorded the training and experience of dedicated law enforcement officers . . . [and] the furtherance of crime prevention must not be discouraged by the hypertechnical application of legal standards."[62] The Tenth Circuit similarly noted that courts should "avoid unrealistic second-guessing of police officers' decisions and . . . accord appropriate deference to the ability of a trained law enforcement officer to distinguish between innocent and suspicious actions."[63] The upshot is that litigants generally find it difficult to convince a court that a *Terry* stop was unjustified, except in

[59] *See, e.g.*, 391 U.S. at 4 (noting police officer's 39 years of experience and his developed habits of observation in crediting officer's statement that defendant "didn't look right to me").
[60] 449 U.S. at 418.
[61] *Id.*
[62] State v. Glover, No. 83-333, 1984 WL 7849, at *3 (Ohio Ct. App. Apr. 20, 1984). *See also* United States v. Johnson, 364 F.3d 1185, 1189-90 (10th Cir. 2004) (overturning lower court's order granting motion to suppress based on finding that district court "impermissibly evaluate[d] and reject[ed] each factor [giving rise to reasonable suspicion] in isolation and failed to accord proper deference to the judgment of an experienced officer").
[63] United States v. Villa-Chaparro, 115 F.3d 797, 801 (10th Cir. 1997)

those uncommon circumstances where an officer openly admits to having no particularized suspicion or to relying solely on suspect criteria.

B. Equal Protection

Contesting the existence of reasonable articulable suspicion is a roundabout way of challenging police discrimination. What about a more direct attack: a straightforward claim that police stops are invalid if they are racially discriminatory, regardless of whether they are supported by reasonable articulable suspicion? The Supreme Court largely blocked that line of attack, at least under the Fourth Amendment, in its 1996 decision in *Whren v. United States*.[64] The Court held in *Whren* that an officer's subjective motivation in executing a stop was wholly irrelevant to the question of the stop's constitutionality under the Fourth Amendment.[65] In fact, an officer's reasons for making a stop *could not* invalidate that stop where the circumstances, seen objectively, justified his action.[66] Accordingly, where an officer pulled over two black motorists who had idled at a stop sign for more than twenty seconds and failed to signal before making a turn, the Court found that the stop was reasonable, despite the possibility that the traffic stop was a pretext for a search predicated on racial bias; all that mattered was that the defendants' behavior gave rise to probable cause that they had violated the traffic code.[67] In reaching that decision, the Court largely foreclosed the Fourth Amendment as an avenue for seeking relief on claims of racial profiling, stating that it was "unwilling to entertain Fourth

[64] 517 U.S. 806.
[65] *See id.* at 812.
[66] *Id.*
[67] *Id*. at 819.

Amendment challenges based on the actual motivations of individual officers."[68] As for defendants who wished to raise claims of racial profiling, it continued, "the constitutional basis for objecting to intentionally discriminatory application of laws is the Equal Protection Clause, not the Fourth Amendment."[69]

The jurisprudence on equal protection, however, presents its own challenges, given the requirement of proof of discriminatory intention. Under the Supreme Court's 1976 decision in *Washington v. Davis*, an Equal Protection claim must allege both disparate impact and discriminatory intent.[70] Without the latter, "official action will not be held unconstitutional solely because it results in a racially disproportionate impact."[71] Under this framework, the Court in *Hunter v. Underwood* found that a provision in the Alabama Constitution disenfranchising persons convicted of crimes involving moral turpitude violated the Fourteenth Amendment: the provision was motivated by racial animus—the president of the constitutional convention explicitly acknowledged the provision's goal as "white supremacy"—and the effect was indisputably disproportionate—"even the most modest estimates" showed that blacks were at least 1.7 times as likely as whites to suffer disenfranchisement under the provision.[72]

---

[68] *Id*. at 813.
[69] *Id*. *Whren* involved a full-fledged arrest based on probable cause, not a brief detention based on reasonable suspicion, but the Supreme Court later made clear—if there were any doubt—that this was of no moment; *Whren* "swe[eps] broadly to reject inquiries into motive generally" in Fourth Amendment analysis, "outside the narrow context of special needs and administrative inspections." Ashcroft v. al-Kidd, 563 U.S. 731, ___; 113 S. Ct. 2074, 2081 (2011).
[70] Washington v. Davis, 426 U.S. 229 (1976)
[71] Village of Arlington Heights v. Metropolitan Housing Development Corp., 429 U.S. 252, 264-65 (1977).
[72] 471 U.S. 222, 227, 229 (1985).

Absent such explicit proof of racial animus, however, class action plaintiffs or criminal defendants seeking to challenge a policy with a disparate impact under the Equal Protection Clause have been largely unsuccessful. The Court's subsequent decisions in the context of selective prosecution challenges affirmed the rigorous standard for Equal Protection challenges. In *Wayte v. United States*, it reaffirmed that "discriminatory intent" implies more than just *awareness* of disparate impact.[73] And in *United States v. Armstrong*, it clarified that, to obtain discovery, criminal defendants must show that similarly situated suspects of a different race or group violated the law and were not prosecuted.[74] This requirement is rigorous: demanding that defendants put forth some evidence tending to show that minorities were singled out *before* obtaining discovery has thwarted the vast majority of such challenges.[75] In *Armstrong* itself, for example, the Court denied the criminal defendant's request for discovery where the defendant showed that all of the prosecutions for crack cocaine were of black defendants, despite anecdotal evidence that blacks and whites used crack cocaine at similar rates.[76]

Because law enforcement agencies do not make or keep records on individuals they do *not* stop, some lower courts have allowed defendants to present statistical evidence of disparate

[73] 470 U.S. 598, 610 (1985) (rejecting criminal defendant's selective prosecution claim where defendant presented evidence that government was aware that enforcement policy would lead to greater prosecutions of particular group and emphasizing that "discriminatory purpose. . . implies more than . . . intent as awareness of consequences").
[74] 517 U.S. 456, 458 (1996). The same standard applies to selective enforcement claims. *See, e.g.,* White v. Williams, 179 F. Supp. 2d 405, 418 at n. 5 (D.N.J. 2002).
[75] *See, e.g.*, United States v. Arenas-Ortiz, 339 F.3d 1066 (9th Cir. 2003); United States v. Hendrickson, 664 F. Supp. 2d 793, 799 (E.D. Mich. 2009); *see generally* Marc Price Wolf, *Proving Race Discrimination in Criminal Cases Using Statistical Evidence*, 4 HASTINGS RACE & POVERTY L.J. 395, 416 (2007).
[76] *Id*. at 459.

impact as an acceptable substitute to the similarly situated showing.[77] Although this more flexible standard paves the way for future challenges in the stop-and-frisk arena, plaintiffs and criminal defendants still bear the burden of proving discriminatory intent—and it is rare that they will have evidence that a police department has intentionally set out to discriminate against minority communities.[78] Discriminatory intent may be inferred from the fact that a law bears more heavily on one race than another,[79] but such impact must be "total or seriously disproportionate" to satisfy the requirement.[80] This was the situation in the 1886 case of *Yick Wo v. Hopkins,* where an ordinance prohibiting laundry operations in wood structures except by permit was enforced only against Chinese residents; *every* Chinese resident was denied a permit, while all non-Chinese residents but one were granted permits.[81] But absent such unusually stark patterns, the judiciary has been reluctant to infer discriminatory intent from statistical evidence of discriminatory impact alone—an attitude that has doomed the vast majority of cases brought under claims of disparate impact, including those brought to combat police discrimination.

C. Statistical Proof of Discriminatory Intent

For the most part, claims of discrimination in policing have generally been unsuccessful in court even when bolstered by statistics. Much of the reason is the Supreme Court's 1987

[77] *See, e.g.*, United States v. Mesa-Roche, 288 F. Supp. 2d 1172, 1186-87 (D. Kan. 2003).
[78] *Floyd v. City of New York* proved an anomaly in this regard; the plaintiffs in that case actually obtained secretly recorded conversations in which police supervisors explicitly urged officers to target black men and in which the Police Commissioner expressed his desire to instill fear in young blacks and Hispanics. 959 F.Supp.2d at 603-05. These uncommon facts helped Judge Scheindlin find the requisite discriminatory intent to satisfy plaintiffs' equal protection claim. *Id*.
[79] *Id*. at 242.
[80] *Id*. at 242.
[81] 118 U.S. 356 (1886).

decision in *McCleskey v. Kemp*,[82] which set sharp limits on the use of statistical evidence to prove unconstitutional discrimination in the criminal justice system. In *McCleskey*, the Court rejected the Equal Protection claim of a criminal defendant who put forward statistical evidence of the racially disproportionate impact of the death penalty. Although the Court acknowledged the significance of the studies presented, conceding at least that the evidence "indicate[d] a discrepancy [in sentencing] that appears to correlate with race," it emphasized that statistical proof must present a "stark" pattern to be accepted as the sole proof of discriminatory intent, and found that the statistical findings were not stark enough to make this inference.[83] This was in spite of the defendant's sophisticated statistical analysis indicating that defendants charged with killing white victims were 4.3 times as likely to receive a death sentence as defendants charged with killing blacks, and that blacks who killed whites were sentenced to death at more than seven times the rate of whites who killed blacks.[84] The Court explained its hesitation to view such findings as dispositive by emphasizing that although the evidence did suggest racial disparities, the defendant had failed to show specifically that the decision-makers in *his* case—which included the prosecutor, the judge, and the individual jury members—had acted with discriminatory purpose.[85] Without such a showing, the Court was unwilling to overturn the death penalty in the defendant's case, noting its concern that such a holding would extend to all capital cases where the victim was white and the defendant black "without regard to the facts of a particular case."[86]

[82] 481 U.S. 279.
[83] *Id*. at 292-93, 312.
[84] *Id*. at 287, 327.
[85] *Id*. at 292-93.
[86] *Id*.

With few exceptions, lower court decisions addressing statistical proof of police discrimination echo *McCleskey*'s reluctance to use statistical evidence to infer discriminatory intent. For example, in *United States v. Mesa-Roche*,[87] the court concluded the defendant had made a "strong showing" that the Kansas state trooper who pulled him over and arrested him discriminated against Hispanic motorists: Hispanics constituted a much higher percentage of the motorists stopped by that officer than of motorists stopped by other officers in the same department.[88] Nonetheless the court rejected the defendant's selective enforcement claim without allowing discovery, because the defendant had "scant direct evidence specific to his case that would support an inference that racial considerations played a part in *his* stop."[89] The same trial judge reached the same conclusion in *United States v. Duque-Nava*,[90] involving a similar claim about a different arresting officer.

The result was identical in *Chavez v. Illinois State Police*, a class action lawsuit alleging that the Illinois State Police was discriminatorily stopping, detaining, and searching black motorists. The Seventh Circuit found in *Chavez* that the plaintiffs could offer no evidence *specific to their case* to support the inference that racial considerations played a part in their stops, despite their showing of a pattern of disproportionate police stops of minorities.[91] In so doing, the *Chavez* court reiterated *McCleskey*'s holding that a litigant must prove discriminatory

[87] 288 F. Supp. 2d 1172 (D. Kan. 2003)
[88] *Id.* at 1190.
[89] *Id*. at 1992 (emphasis added).
[90] 315 F. Supp. 2d 1144, 1163 (D. Kan. 2004).
[91] 251 F.3d 612, 645 (7th Cir. 2001)

intent by showing that the alleged discriminators "selected or reaffirmed a particular course of action at least in part because of its adverse effects upon an identifiable group."[92]

Similarly, in *Anderson v. Cornejo*, the Seventh Circuit rejected a class action claim brought against supervising officers by black women searched at O'Hare International Airport in Chicago.[93] The plaintiffs offered "hit rate" evidence that contraband was only found on 27.6% of black women who were strip-searched, compared to 61.6% of black men, 58.8% of Hispanic men, and 45.7% of Hispanic women; these results, they argued, indicated that customs inspectors searched black women with less by way of suspicion than they required before searching Hispanic or black men (although the results also indicated that black women seem to have been treated similarly to both white men and white women).[94] Echoing the familiar refrain, Judge Easterbrook reasoned that "these statistics show disparate impact, not disparate treatment, and the equal protection guarantee is concerned only with the latter."[95]

Another example: In *United States v. Avery*, the Sixth Circuit rejected a selective enforcement claim based on evidence, similar to the proof in *Anderson*, that an airport drug interdiction unit stopped blacks disproportionately.[96] The evidence in *Avery* suggested that blacks were a majority of the travelers stopped by narcotics officers at the Cincinnati airport, although they were a minority of the passengers passing through the airport.[97] As in *McCleskey*, the court cited *Yick W*o to emphasize that statistical evidence must show striking disparities to

[92] *Id*. at 645 (internal quotation marks omitted).
[93] 355 F.3d 1021 (7th Cir. 2004)
[94] *Id*. at 1023.
[95] *Id*. at 1024.
[96] 137 F.3d 343 (1997).
[97] *Id*. at 348.

satisfy the intent requirement, and held that the defendant had not so shown. Furthermore, even if the data suggested general disparities in police stops, the defendant had not shown that the officers had been racially motivated in stopping *him*; the court accepted the officers' explanation that they had reasonable, particularized suspicion where the defendant was walking in a hurry through the airport wearing short sleeves in December.[98]

Finally, and most recently, in *United States v. Johnson*, a federal district court in North Carolina dismissed equal protection claims brought by the U.S. Department of Justice against a county sheriff based on statistical evidence.[99] The government had presented proof that Hispanic drivers were stopped and cited at a much higher rate than the rate at which they violated traffic laws, and that approximately 6% of searches of stopped Hispanics uncovered drugs, compared to 30% of searches of stopped non-Hispanics.[100] This evidence, the government contended, suggested that a lower threshold of suspicion or probable cause was applied to Hispanics.[101] The court disagreed that such evidence could make out a Fourteenth Amendment claim, repeating that in "situations similar to the present case," the rule remained that "statistics may not be the sole proof of a constitutional violation."[102]

[98] *Id.* at 358.
[99] United States v. Johnson, 2015 WL 4715312 (M.D.N.C., Aug. 7, 2015, 1:12CV1349)
[100] *Id*. at *26.
[101] *Id*. at *28.
[102] *Id.* at *71, fn. 88.

The bottom line is that unless officers admit to racial animus, or perhaps if the disparate impact is as striking as in *Yick Wo*, an equal protection challenge is likely to fail for lack of proof that the police *in this particular case or set of cases* acted with discriminatory purpose.[103]

This is not to say that such challenges have never succeeded. For example, in *State v. Soto*, a New Jersey state court found that a group of seventeen black defendants subjected to police highway stops had made out a *prima facie* case of selective enforcement on the basis of data showing that black motorists were stopped disproportionately.[104] But the defendants in *Soto*, who had been stopped for speeding, were able to compare the racial distribution of motorists stopped for speeding in New Jersey not just with the racial distribution of New Jersey residents but with the racial distribution of speeders on the highway where the defendants had been stopped; they were helped by researchers who had calibrated their speedometers and recorded the number of vehicles overtaking them. (The results showed that approximately 15% of all speed limit violators were black, compared to 46% of all motorists stopped for speeding.[105]) In most cases, statistics on violators—as opposed to the general population and the individuals who are stopped by the police—are difficult to come by, so efforts to rely on *Soto* in other cases have generally failed.[106] And even where plaintiffs have the resources to mimic the methodology in *Soto*—as in *Johnson*, where researchers observed all drivers on selected roads, their ethnicity,

---

[103] Sometimes the bar may be even higher. For example, at least one court has ruled that the Equal Protection Clause is not violated by racial bias in the application of the death penalty if the bias is unconscious; such bias, the court reasoned, "stems from a psychological phenomenon rather than some flaw created by the legal system and its procedures." In re Death Penalty Claims, 2013 WL 5879422, at *24 (Conn. Super. Oct. 11, 2013) (unpublished opinion).
[104] 324 N.J. Super. 66 (Ch. Div. 1996)
[105] *Id*. at 70.
[106] *See* United States v. Barlow, 310 U.S. 1007 (7th Cir. 2002); Commonwealth v. Lora, 886 N.E.2d 688 (Mass. 2008).

and whether they were violating a North Carolina traffic law—courts remain skeptical that the comparison group, or "benchmark," is sufficiently rigorous. In *Johnson* itself, the court rejected the value of the researchers' observations because "no control, standard, or description was used to identify Hispanics in his study"; the fact that "the surveyor observing virtually all of the drivers simply identified people as Hispanic if he thought they 'appeared to be' Hispanic" rendered the study unreliable.[107]

Still, it is noteworthy that recent cases rejecting statistical proof of discriminatory policing tend to suggest that stronger statistics might have led to a different result. In *Chavez*, for example, the court took great care to explain why the specific statistics presented were not reliable evidence of the actual racial distribution of stops.[108] First, the court pointed out that the plaintiffs seemed to have cherry-picked a potentially non-random and statistically insignificant sample of police stops—which were themselves only selectively recorded—for analysis.[109] Second, the court took issue with the plaintiffs' population benchmarks. The plaintiffs had compared the proportion of minorities stopped by the police with their share of the state population and the state's motorists, but the population information was based on out-of-date census reports, and the data on motorists came from a survey that had included very few blacks and Hispanics.[110] Had the analysis been based on better data, the court suggested that the outcome for the plaintiffs may have been different.[111]

[107] 2015 WL 4715312, at *48.
[108] *See* 251 F.3d at 640-44.
[109] *See id*. at 641-43.
[110] *See id*. at 644.
[111] *See id*. at 642.

Similarly, while Judge Easterbrook pointed out weaknesses in the *Anderson v. Cornejo* plaintiffs' data—the data reported were not sampled from the same year as when the alleged violations occurred, thus making it difficult to extrapolate results; they were based on national estimates of stops and searches, rather than from the airport at issue; and there was no control for other non-discriminatory characteristics that could explain the disparity in hit rates—he did not reject the hit rate method itself as a way of proving discriminatory effect.[112] Likewise, the *Johnson* court emphasized that the studies were insufficient to prove intent "because of the multiple deficiencies noted" with the *particular* statistical evidence presented: for example, the hit rate analysis did not provide controls for the type of search—like consent, search warrant, or protective frisk—despite the fact that consent searches may have explained "a significant portion of the searches" and thus accounted for the disparity between races.[113] In *Avery*, the Sixth Circuit noted that the defendant's data appeared to be a skewed sample of only about half of all narcotics stops at the Cincinnati airport in the relevant time period[114]; the court was explicit that "valid relevant statistical evidence of disparate impact" *could* prove discriminatory intent.[115]

The federal trial judge in *Mesa-Roche* went further, describing the kinds of statistics that might well suffice to demonstrate discrimination. She noted that "a study would be well on the road towards an identification of racial motive or intent using a valid scientific methodology, if the evaluative process of the law enforcement officer [were] observed in process, and data [were] collected on the various factors used with respect to each stop . . . . The identification of appropriate, race-neutral, sound and effective factors could then be used to evaluate an individual

[112] 355 F.3d at 1023-25.
[113] 2015 WL 4715312, at *81.
[114] 137 F.3d at 357.
[115] *Id*. at 355.

officer's decision making, through logistic regression analysis."[116] As we now will explain, this is precisely the kind of study that big data is making increasingly feasible.

II. Police Discrimination and Big Data

Despite longstanding concerns about discriminatory policing—and, in particular, about racial bias in stop-and-frisk practices—litigants have had only spotty success in mounting challenges to racially discriminatory *Terry* stops. The requirement of articulable suspicion has proven of sharply limited value in ferreting out discriminatory motive, partly because the courts have tended to credit any facially neutral explanation an officer provides for a stop-and-frisk, and partly because the Supreme Court has declared an officer's actual motivations immaterial so long as there were objectively reasonable justifications for the intrusion. Equal protection challenges, meanwhile, typically have foundered for lack of compelling proof of discriminatory intent. The advent of big data has the potential to change this situation. Indeed, it has already begun to do so.

A. The Growing Availability of Data on Police Stops

Thirty years after *Terry v. Ohio*, the New York Police Department was still the only major police agency in the U.S. regularly collecting comprehensive data on the stop-and-frisk practice.[117] But the past fifteen years have seen a steady shift toward greater data collection on such practices.[118] As a result of consent decrees, class action settlements, state statutes, and

[116] 288 F. Supp. 2d at 1196-97.
[117] *See* Harris, *supra* note 23, at 870-71.
[118] *See id.* at 856.

police department agreements with local city councils, police officers around the country are now required to record data on every stop.[119]

In New York, the police department's practice of collecting data on stops pre-dates *Terry*; beginning in 1964, the NYPD required officers to record stops on a standard form called the UF-250.[120] However, it was not until 1999 that records from this data collection effort were released to the public, when New York's Attorney General ordered the NYPD to turn over two years of data following weeks of protests in response to the NYPD's killing of an unarmed black man.[121] Although the NYPD has continued to record information about stops, the department also continued to resist publishing their records until 2008, when a state judge ordered them to make their electronic database public in response to a lawsuit filed by the New York Civil Liberties Union.[122] More recently, Judge Scheindlin's opinion in *Floyd v. City of New York* reinforced the NYPD's mandate to continue collecting (and publishing) data, and imposed a proactive duty on the department to meaningfully audit their records to monitor the constitutionality of stops.[123]

Police departments in many metropolitan cities in the United States have followed New York's lead, if often unwillingly. A survey of forty-four of the largest police departments in the country revealed that over half—twenty-three of them—required officers to collect some specific

---

[119] *See id*. at 869-71. Some of the most important settlements have come in investigations or lawsuits the federal Department of Justice has instituted against local police departments under 42 U.S.C. § 14141. *See supra* note 41; Fan, *supra* note 3, at 127-28.
[120] U.S. COMM'N ON CIVIL RIGHTS, POLICE PRACTICES AND CIVIL RIGHTS IN NEW YORK CITY, ch. 5, n. 62 (2000), *available at* < www.usccr.gov/pubs/nypolice/ch5.htm>.
[121] N.Y. STATE OFFICE OF THE ATTORNEY GEN'L, THE NEW YORK CITY POLICE DEPARTMENT'S "STOP & FRISK" PRACTICES (1999).
[122] Christine Hauser, *Police Told To Give Street-Stop Data*, N.Y. Times, May 31, 2008, *available at* <www.nyt.com>.
[123] 959 F. Supp. 2d at 608.

data on each stop and frisk conducted, while eighteen did not require any data collection, and three maintained records, but made data collection discretionary.[124] Of those departments that mandated collection, a few—like Los Angeles and Cincinnati—did so as the result of consent decrees after the Department of Justice initiated investigations into the cities' police practices.[125] Others, like Philadelphia and Oakland, began collecting data as the result of lawsuit settlements.[126]

The limitations of the available data are still daunting. The data collected is not uniform across departments; for example, of those departments that collect data, Oakland and Los Angeles do not systematically or consistently record the race of the suspect,[127] and Chicago does not record stops that lead to tickets or arrests, or record whether an officer also conducted a frisk pursuant to a stop.[128] Even in New York, which leads the nation in documenting these encounters, a federal monitor's report to the court found that in over one-quarter of stops in 2015, officers failed to properly document the suspicion that prompted them to stop someone.[129] More troubling, for those departments keeping records outside of New York, the data is often

[124] Harris, *supra* note 23.
[125] *See id.* at 871; United States v. City of Los Angeles, No. 00-11769 GAF, Consent Decree, § 3 (C.D. Cal. 2001) (requiring the Los Angeles Police Department to require its officers "to complete a written or electronic report each time an officer conducts a pedestrian stop").
[126] *See, e.g.,* Harris, *supra* note 23, at 871-72; Bailey v. City of Philadelphia, No. 10-5952, Settlement Agreement, Class Certification, and Consent Decree, at 3 (E.D. Pa 2011).
[127] *See* Harris, *supra* note 23, Appendix A.
[128] *See* ACLU OF ILLINOIS, STOP AND FRISK IN CHICAGO (2015).
[129] Submission of Second Report of Independent Monitor, Floyd v. City of New York, 08-CV-1034, No. 523 (February 16, 2016).

inaccessible for review outside of the police department.[130] Furthermore, there is not yet any national database on police stops.[131]

The clear trend, though, is towards more and better data collection. In some cities, the police departments themselves are initiating analysis into their search patterns.[132] The Center for Policing Equity, a research consortium founded in 2008 to foster collaboration between law enforcement agencies and researchers, is currently working on the first database tracking national statistics on police stops, together with the Department of Justice.[133] As part of the project, the researchers are partnering with police departments to improve their recordkeeping, and standardizing data across departments.[134] So far, more than fifty police departments have expressed interest in participating in the compilation of data.[135] A separate project at Stanford, in which two of us are involved, is compiling a comprehensive, systematically organized database of more than 100 million traffic stops across the United States.[136] Meanwhile, other cities

---

[130] Letter Harvey Grossman, Legal Dir., ACLU of Illinois, to Chicago Mayor Rahm Emanuel, Jan. 15, 2013, *available at* <www.aclu-il.org/wp-content/uploads/2013/ 01/Letter-to-Emanuel-Patton-McCarthy-1-15-13.pdf>.
[131] Arun Rath, Taking On Racial Profiling With Data, NPR, Dec. 14, 2014, *available at* <www.wbur.org/npr/370792960/taking-on-racial-profiling-with-data>; Phillip Atiba Goff, *Police Behavior Database: Why One Doesn't Exist and Why One Soon Will*, The New Republic, August 14, 2014, *available at* <www.newrepublic.com>.
[132] Richard A. Oppel, Jr., *Activists Wield Search Data to Challenge and Change Police Policy*, N.Y. TIMES, Nov. 20, 2014, *available at* <www.nyt.com>.
[133] Center for Policing Equity, Justice Database, *available at* <policingequity.org>; *see also* Center for Policing Equity, Press Release: Nation's First Police Profiling Database Awarded Grant by NSF, Nov. 7, 2013, *available at* <policingequity.org>; U.S. Dep't of Justice, Office of Public Affairs, Justice Department Announces National Effort to Build Trust Between Law Enforcement and the Communities They Serve, Sept. 18, 2014, *available at* <www.justice.gov/opa/pr/ justice-department-announces-national-effort-build-trust-between-law-enforcement-and>.
[134] *Id.*
[135] *See* Oppel, *supra* note 132.
[136] *See* Brady Dale, *Stanford Traffic Stops Database Will Let Public Analyze Racial Profiling*,

continue to independently undertake efforts to engage in systematic data collection in reaction to the legal challenges to stop-and-frisk in New York.[137]

These developments suggest we are entering a new era for raising statistical claims of reasonableness and discrimination. The increasing availability of data opens the way to more rigorous analysis of police practices. The decision in *Floyd v. City of New York* is particularly instructive in this regard, providing a preview of the way statistical findings of disparities in policing patterns can bolster challenges to the constitutionality of the program as a whole—as well as the extent to which courts continue to rely on non-statistical evidence in assessing discriminatory intent, and continue to resist the use of statistical evidence when assessing the reasonableness of a particular stop of a particular suspect.

B. *Floyd v. City of New York*

Floyd v. City of New York represented a landmark victory for class action plaintiffs challenging their stops under the Fourth and Fourteenth Amendments. In that decision, District Court Judge Shira Scheindlin expanded their protection to encompass claims brought largely on the basis of statistical evidence, although she stopped short of holding that statistical evidence alone would be sufficient.

OBSERVER, Feb. 1, 2016, *available at* <www.observer.com>.

[137] *See, e.g.*, San Diego Union-Tribune, *San Diego Police to Cull Racial Data During Stops*, Jan. 22, 2014, *available at* <www.utsandiego.com>.

With regard to the plaintiffs' Fourth Amendment claim, Judge Scheindlin found violations based on a combination of qualitative and quantitative evidence. First, she considered the rationales for stops provided on UF-250s, the forms that officers are required to prepare after each stop, finding that stops made on the basis of certain factors, like "furtive movements," gave officers too much discretion.[138] This was because such factors were "vague and subjective" and could be affected by unconscious racial biases.[139] Similarly, she found that the "high crime area" descriptor was of "questionable value," given that the analysis of officers' forms revealed that they often used it to describe entire boroughs, and that the instances in which they checked the box did not correlate to actual crime rates in different areas.[140] Finally, she relied on a version of hit rate analysis to find that reliance on these "vague" factors led to ineffective stops, noting that both "furtive movements" and "high crime area" were negatively correlated with arrest rates. That is, stops were more likely to result in arrest where such factors were *not* checked,[141] suggesting that officers had a stronger basis for effecting an arrest when they relied on something other than those two factors, and that their suspicions were less well-founded where they did not. Similarly, the fact that only 1.5% of frisks revealed weapons signaled to her that too many of these intrusions were unfounded and thus unconstitutional.[142] However, while these considerations led her to the conclusion that the plaintiffs had established *some* Fourth Amendment violations, she emphasized that determining exactly which stops had been unconstitutional "will almost certainly never be known," beyond ascertaining a "rough

[138] 959 F. Supp. 2d at 575, 579-80, 614.
[139] *Id*. at 579.
[140] *Id*. at 579, 581.
[141] *Id*. at 575. Specifically, stops were 22% more likely to result in arrest if "High Crime Area" was not checked, and 18% more likely to result in arrest if "Furtive Movements" was not checked.
[142] *Id*. at 660.

minimum" of unconstitutional stops for which officers had listed *no* reason for the stop, or only checked a factor that higher courts had already dismissed as being insufficient to justify a stop.[143]

For her Fourteenth Amendment analysis, Judge Scheindlin relied on benchmark studies to determine whether blacks and Hispanics were stopped disproportionately in relation to their racial composition in census tracts.[144] The findings that blacks and Hispanics were more likely to be stopped than whites, even after controlling for crime rates, racial composition of the area, and various other factors, and that the NYPD carried out more stops in areas with more black and Hispanic residents, again controlling for relevant variables, led her to conclude that the stop-and-frisk policy had a racially disproportionate impact. Together with evidence that the police department acted with discriminatory intent—for example, by explicitly directing officers to target young black and Hispanic men—this finding compelled her conclusion that the policy, as implemented, violated the Fourteenth Amendment.[145]

The decision advanced the jurisprudence on racial profiling, relying as it did on statistics to find the stops and searches unconstitutional. However, the opinion largely maintains the existing frameworks in the Fourth and Fourteenth Amendment contexts. First, it reaffirmed that an officer's stated rationales may continue to lend searches legitimacy. Although Judge Scheindlin criticized the "furtive movement" and "high crime area" categories as giving officers too much discretion, she stopped short of flatly declaring unconstitutional all stops based on such written justification, merely noting that their negative correlation with summons or arrests

[143] *Id*. at 583.
[144] *Id.* at 583-84.
[145] *Id*. at 663.

suggested that these stops were likely unjustified.[146] To the extent that she concluded that any stops or frisks violated the Fourth Amendment, she limited these to stops for which *no* rationale was given,[147] stops for which officers only checked factors that she deemed inherently insufficient,[148] and instances where plaintiffs could put forward detailed and individualized evidence regarding the unreasonable circumstances of their stop or frisk.[149]

This leaves a gap for others seeking to challenge the constitutionality of their stops. While Judge Scheindlin found the general program of stop-and-frisks in New York unconstitutional, she largely refrained from determining exactly in which cases stops were unreasonable, apart from fact-specific inquiries into the nineteen Named Plaintiffs' experiences. Although she viewed the low hit rate for stops based on certain factors like "furtive movement" and "high crime area" as probative—but not dispositive—of their potential unreasonableness, her opinion leaves undisturbed all the stops for which other rationales are listed. Rather, her view that determining which stops are or are not justified depends on "individually analyz[ing] each of those stops"[150] makes it clear that a criminal defendant seeking to challenge his particular stop is back in the traditional regime we discussed earlier.

With regard to the Fourteenth Amendment, Judge Scheindlin relied on the plaintiffs' use of statistics in finding disparate impact, but was careful to separate their statistical evidence of disparities from their ability to show discriminatory intent. Rather, her analysis of the Equal

[146] *Id*. at 582-83.
[147] *Id*. at 579-582.
[148] *Id.*
[149] *Id*. at 625-58.
[150] *Id*. at 578.

Protection claim reaffirmed the Supreme Court's instruction that plaintiffs must show that those responsible for profiling did so "because of, not merely in spite of, its adverse effects upon the profiled racial groups."[151] To meet this requirement, Judge Scheindlin drew upon the police department's explicit references to race and admissions that the department "targeted young blacks and Hispanics for stops."[152] Thus, her analysis did not attempt to suggest that statistical proof alone was enough to infer discriminatory intent, even given the "stark racial disparities"[153] she found in New York's stop-and-frisk regime. Despite the path-breaking nature of the *Floyd* decision, it therefore left open the possibility that future equal protection challenges to discriminatory policing might remain largely dependent on official statements explicitly demonstrating the existence of impermissible motives.

III. Beyond *Floyd*: Stop-Level Hit Rate Analysis

Floyd v. City of New York dramatically demonstrated the growing power of big data to help detect illegal discrimination in a stop-and-frisk program, but new techniques of statistical analysis make big data an even stronger tool for these purposes. We will illustrate this point by describing a technique that two of us, with a third co-author, have developed for using police data to calculate stop-level hit rates (SHRs)—numerical estimates of the likelihood that the suspicion motivating a *Terry* stop will turn out to be correct.[154] We have applied this technique to NYPD stops for suspicion of criminal possession of a weapon (CPW), but it could be applied in

[151] *Id*. at 662 (internal quotation marks omitted).
[152] *Id*. at 663.
[153] *Id*. at 662.
[154] *See* Goel, Rao & Shroff, *supra* note 25.

other cities that are now collecting systematic data on *Terry* stops, and for reasons other than CPW as well.[155]

The basic idea of SHR analysis is to use a statistical model to calculate the *ex ante* likelihood, based on the information available to the officer, that a *Terry* stop will be "successful"—i.e., will result in finding what the officers suspect they will find. To compute these SHRs, we used the information recorded on the NYPD's UF-250 reports, but only the information that would have been available at the time the officer made the decision to conduct a stop-and-frisk. Specifically, we first fitted a logistic regression model[156] to the 472,344 CPW stops recorded by NYPD officers from 2008 through 2010, defining "success" as the discovery of a weapon.[157] We included in the model variables for demographic information about the suspect (sex, race, age, height, weight, and build); location of the stop (precinct; inside or outside; and on public transit, in public housing, or neither), date and time of the stop (year, month, day of week, and time of day); the recorded reasons for the stop (e.g., "furtive movements" or "high crime area"); whether the stop was the result of a radio run; whether the

[155] For example, we have applied the same technique for stops motivated by suspicion of drug sale or possession. *See* Goel, Rao & Shroff, *supra* note 25, at __.

[156] Logistic regression is a widely used statistical technique for estimating the likelihood of a binary outcome (e.g., whether or not an individual will be found to have a weapon) based on one or more predictor variables (e.g., the time of day, location, and suspect characteristics). To construct (or "fit") the model, one uses a dataset in which both the predictors and the outcomes are recorded for a representative sample of cases; in our setting, we use data from hundreds of thousands of stops recorded between 2008 and 2010. Then, given only values of the predictor variables (but not the outcome itself) for a new case, the fitted logistic regression model outputs the probability, based on the provided information, of seeing each of the two possible outcomes (e.g., weapon or no weapon) in that case.

[157] We defined success as finding a weapon because these stops were explicitly justified by suspicion of CPW. For other kinds of stops, a different measure of success could be used—e.g., discovery of other kinds of evidence or contraband, or an arrest or a summons (the criterion used for calculating "hit" rates in *Floyd*).

officer was in uniform; how long the officer observed the suspect before initiating the stop; and the "local hit rate" of stops at that location. The local hit rate at a given location is the weighted percentage of CPW stops during the previous year for which a weapon was recovered. Stops close to the given location receive higher weights than stops that are further away.[158] The model also included all pairwise interactions between these variables, resulting in 7,705 predictive features. We then used the fitted model to estimate the *ex ante* likelihood that each of 288,158 stops in the *following* two years—2011 and 2012—would turn up a weapon. The SHR model is thus very much a product of the age of big data. It could not have been developed or implemented without extensive stop-level data, and it is designed to take full advantage of this detailed information.

The results produced by the SHR method are dramatic. First, the model turns out to be highly accurate. To evaluate the model, we selected random pairs of cases from among the 2011 and 2012 stops where a weapon was ultimately found in exactly one stop of the pair. Presented only with the stop-level predictors (and not the outcomes), a completely uninformative model would do no better than chance at determining in which one of the two stops a weapon was found. In contrast, we found that our SHR model correctly picked out the stop with the weapon 83% of the time, indicating good predictive performance.[159] We further checked the "calibration" of the model. We found that among cases in which the model said an outcome would occur with probability *p*, the outcome in fact occurred approximately *p* percent of the time. For example, in

[158] Of course, this reflects a different construction of "high crime" neighborhoods than that indicated in the UF-250s; as Judge Scheindlin noted, officers used the "high crime" factor on the form to refer to entire boroughs, regardless of empirical data.
[159] This performance metric is known as the AUC ("area under the curve"), and it is popular in the computer science community for evaluating logistic regression models.

the 7,310 cases where the model determined the likelihood of finding a weapon was 5%, a weapon was ultimately found in 367 (or 5%) of those instances. Finally, since a suspect's height, weight, and age can only be approximated by the officer before the stop, we confirmed the model is largely robust to reasonable errors in these terms. In particular, if we assume officers estimate height, weight, and age with average errors of 2 inches, 10 pounds, and 5 years, respectively, the SHR changes on average by only one-tenth of one-percent. The totality of evidence thus suggests our modeling framework produces accurate and robust estimates.[160]

One implication of this result is that the UF-250 forms indeed contain meaningful information, which in turn reaffirms that documenting stops is a valuable practice for police departments. Because officers complete UF-250 forms after carrying out a *Terry* stop, not before, there is clearly a risk that the forms do not identify the actual reasons for the stop but instead a *post hoc* justification.[161] Moreover, there is evidence, credited by Judge Scheindlin in *Floyd*, that officers follow "scripts of suspicion" when recording the basis for a stop.[162] We did not, however, find any indication of widespread strategizing by officers. For example, the vast majority of CPW stops did not result in the discovery of a weapon, and non-CPW stops often

[160] *See* Goel, Rao & Shroff, *supra* note 25, at 10. The model works well at determining the likelihood that people who were in fact stopped by the police would be found to have a weapon. It does not, however, automatically extend to people who were *not* stopped. One cannot and should not mechanically apply the model to estimate the likelihood that an arbitrary person on the street has a weapon, since those who were not stopped might be qualitatively different from those who were stopped, in a manner that is not captured by the model. Similarly, the model only yields estimates for the current practice of stop-and-frisk. If officers change their stop procedures, the model may no longer be reliable, and would need to be re-fit to new data.

[161] In some instances, that plainly occurs. *See infra* note 167.

[162] Jeffrey Fagan & Amanda Geller, *Following the Script: Narratives of Suspicion in Terry Stops in Street Policing*, 82 U. CHI. L. REV. 51, 55 (2015); *accord*, *Floyd*, 959 F. Supp. 2d at 581.

did, so officers clearly are not classifying a stop as CPW if and only if they find a weapon. Nor are there obvious reasons for officers systematically to favor certain stop factors after stops that are "successful" and other factors after stops that do not lead to the discovery of evidence.[163]

Second, the SHR method suggests that the justification for many *Terry* stops has been quite thin. Figure 1, for example, shows the distribution of success likelihoods for CPW stops the NYPD conducted in 2011 and 2012. For every threshold between 0% and 10%, the plot shows the fraction of stops with *ex ante* likelihood of success below that threshold. The big news is the steep slope of the left-hand side of the plot: 43% of stops had less than a 1% chance of turning up a weapon (typically a knife) and 19% of stops had less than a 0.5% chance.[164]

[163] We cannot rule that possibility out. Perhaps, for example, officers are more likely to claim they saw a "furtive movement," rather than a "suspicious bulge," when no weapon is discovered. They might believe that in the absence of a weapon, a report of a "furtive movement" would be more credible than a claim that the officer saw a "suspicious bulge." But there is no particular reason to believe that NYPD officers thought that way.

[164] *See* Goel, Rao & Shroff, *supra* note 25, at 10.

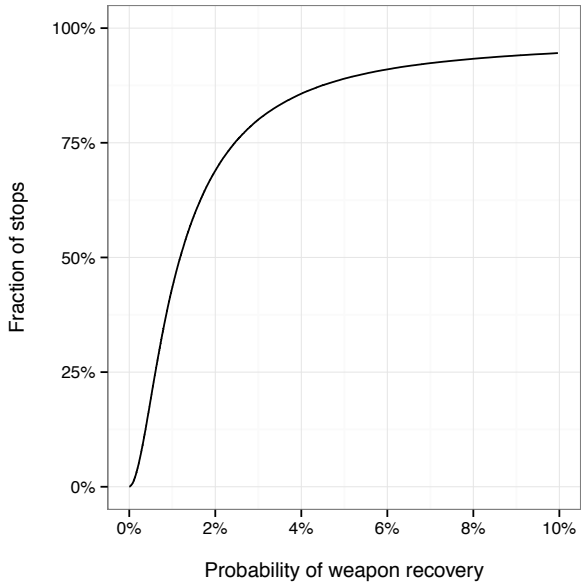

Figure 1. Distribution of the ex ante probability of finding a weapon on a suspect stopped for suspicion of criminal possession of a weapon (CPW), for all such stops between 2011 and 2012. The y-axis indicates the fraction of stops below the x-axis threshold. In particular, 43% of all CPW stops have less than a 1% ex ante chance of turning up a weapon.

Third, the SHR method provides strong, numerical support for the conclusion reached in *Floyd*: that the stop-and-frisk practices of the NYPD discriminated against racial minorities, particularly blacks. Figure 2 disaggregates the success likelihood plots for the NYPD's CPW stops in 2011 and 2012 according to race. The takeaway is that stops of blacks and Hispanics were typically based on less evidence than stops of whites. For example, 49% of blacks and 34% of Hispanics stopped under suspicion of CPW had less than a 1% chance of possessing a

weapon, compared to 19% of whites.[165] Thus, regardless of whether such low hit rate stops violate the Fourth Amendment, they do place a disproportionate burden on minorities.

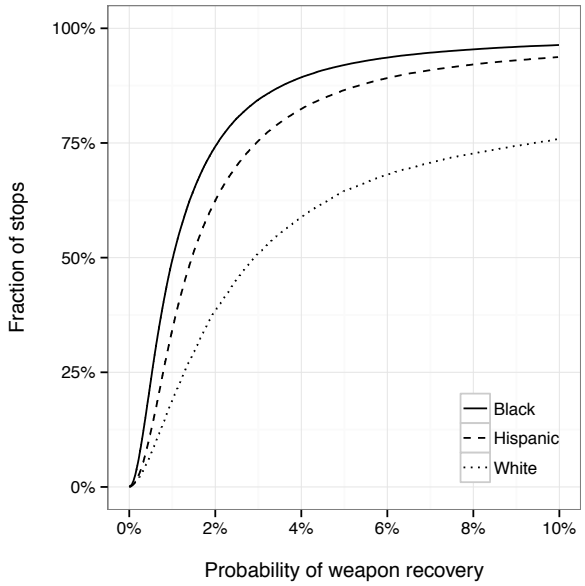

Figure 2. Distribution of the ex ante probability of finding a weapon on a suspect stopped for suspicion of criminal possession of a weapon (CPW) broken down by race, for all such stops between 2011 and 2012. The y-axis indicates the fraction of stops below the x-axis threshold. Blacks and Hispanics are typically stopped on the basis of less evidence than whites.

There could, of course, be race-neutral reasons for stopping minorities based on less evidence. In particular, officers might have a lower bar for stopping individuals in high-crime areas, regardless of race. Since high-crime areas tend to be predominately minority, targeting

[165] *Id*. at 12.

Terry stops in those areas would result in a pattern of minorities, on average, being stopped with less evidence than whites, even in the absence of discrimination.

But even after controlling for location, the SHR method indicates the racial disparities persist. Figure 3 illustrates the analysis. Each point in the plot compares the success rate (i.e., "hit rate") of stops of whites and blacks in a particular location. The success rates vary dramatically across areas, ranging from less than 1% for stops in public housing complexes in Brownsville to more than 30% for transit stops in Midtown Manhattan. The variation suggests that officers do in fact apply different standards for stopping individuals depending on the context. Even within locations, though, the success rate of white stops is higher than that of black stops, indicating that black stops are conducted on the basis of less evidence. And this is true in virtually every location, which is why almost all of the points are above the diagonal line.[166]

[166] The results are similar for Hispanic stops compared with white stops.

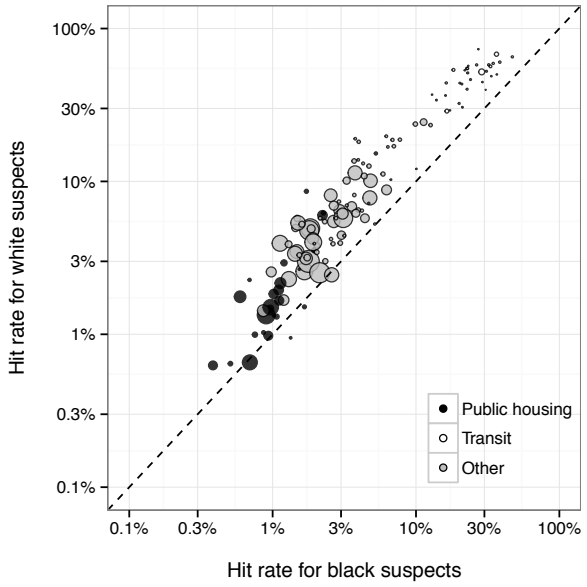

Figure 3. Likelihood of finding a weapon (hit rate) on suspects stopped for suspicion of CPW by geographic area, where we consider only stops of suspects who are black or white. Each circle corresponds to stops within a given precinct in one of three possible location types, indicated by their shading: public housing, public transit, or other (typically street stops). The areas of the circles indicate the number of stops in that location. In nearly every location, the hit rate for black stops is lower than for white stops, indicating that blacks are stopped on the basis of less evidence than whites.

This evidence of discriminatory intent provided by the SHR method is circumstantial; in this respect it is similar to past efforts to use statistics to prove police discrimination. But the SHR method, when used on data as detailed and extensive as the information collected by the NYPD on *Terry* stops, allows for a much more nuanced, and hence significantly stronger, analysis. In particular, the SHR analysis allows researchers to consider and to control for highly

specific stop information, including location, time, suspect characteristics, and the grounds for suspicion identified by the officer.

Fourth, the SHR method not only allows one to estimate the aggregate number of stops that fall below a specified probability threshold, but also yields a quantitative measure of the evidence supporting a stop-and-frisk in each *particular* case, which can in turn be used to determine whether "reasonable articulable suspicion" existed. This use of the SHR technique as a retrospective test of articulable suspicion can be illustrated by using the model constructed from CPW data in New York City to estimate the *ex ante* likelihood of stop success for four of the plaintiffs in *Floyd.* The results are shown in Table 1. For example, SHR analysis indicates that there was a 3% chance that Devin Almonor—a thin, 5 foot, 10 inch 13-year-old black teenager in Harlem who "fits description" and was behaving "furtively"—would be found to have a weapon. In fact, no weapon was found on Almonor, and Judge Scheindlin ruled his stop "unreasonable" under the Fourth Amendment.[167]

[167] 959 F. Supp. 2d at 630. Judge Scheindlin noted that the computerized UF-250 completed the day after Almanor's stop indicated that the officer had seen a suspicious bulge—even though the corresponding box was not checked on the form filled out immediately after the stop, and the officer later testified that he had not, in fact, seen a suspicious bulge. *Id.* at 629-30. The discrepancy underscores the risks inherent in relying on *post hoc* explanations of the factors motivating *Terry* stops. As we have noted, though, we did not see evidence of widespread strategizing by officers in the selection among suspicion factors. *See supra* text accompanying notes 162-163 .

Name	Date & Time	Precinct	Circumstances	Prob.
Devin Almonor	3/20/2010, 8pm (Saturday)	30	Fits description Furtive movements Radio run	3%
Cornelio McDonald	12/18/2009, 1am (Friday)	107	Suspicious bulge Investigation	4%
Nicholas Peart	8/5/2006, 5am (Saturday)	24	Fits description Suspicious bulge Witness report Proximity Radio Run Officer in uniform	1%
David Ourlicht	1/30/2008, 2pm (Wednesday)	107	Suspicious bulge Officer in uniform	4%

Table 1. The ex ante probability of stop success for four of the plaintiffs in Floyd.

There are obvious difficulties in translating the "reasonable articulable suspicion" standard into a probability threshold. The Supreme Court has consistently declined to provide minimum numerical probabilities for either "reasonable articulable suspicion" or "probable cause," and if a line were to be drawn, it is far from clear where it should be located. We will address this problem in more detail below. For now, suffice it to say that if Judge Scheindlin was right that the evidence supporting Almonor's stop fell short of "reasonable articulable suspicion," there is a straightforward argument for saying the same about other *Terry* stops with less than a 3% *ex ante* likelihood of success. As can be seen in Figure 1, that translates to approximately 80% of the CPW stops recorded by the NYPD in 2011 and 2012.

Finally—and more speculatively—SHR analysis can be used not just to assess an officer's decision to conduct a *Terry* stop after the fact, but also to guide that decision in the first place. Because the SHR is calculated from information available to the officer at the time the decision is made to carry out a stop-and-frisk, the method also could be used, in theory, to inform

the stop decision. The logistic regression model fitted to NYPD's UF-250 data is perhaps too complicated for officers to make use of in real time (though hand-held computers could help). We find, however, that a drastically simplified version can preserve much of the predictive value of the full model. Suppose officers calculated a "*Terry* score" for possible subjects of a stop-and-frisk by awarding one point for "suspicious bulge" one point for "sights and sounds of criminal activity," and three points for "suspicious object" (factors that are already recorded on the UF-250 form). And suppose officers were instructed that, in general, they should conduct a stop-and-frisk only when the *Terry* score met or exceeded some threshold—one, two, three, four, or five—with the choice of the threshold dependent on the area in which they were operating. The NYPD's data suggests that this method would allow the police to have recovered half of the weapons they found by conducting just 8% of the stops they actually carried out, and 90% of weapons with 66% of the stops.[168] Notably, we further find that stops with the lowest *Terry* score (i.e., stops based on the least amount of evidence) disproportionately involve minorities. Consequently, by focusing police efforts on those stops most likely to be successful, one can simultaneously increase both the efficiency and racial equity of stop-and-frisk practices.

Such a *Terry* score may seem unrealistic when applied to on-the-spot, highly contextualized decisions by officers acting on their own initiative and applying their own, accumulated knowledge of the areas they patrol. Some *Terry* stops are in fact of that kind. In the main, though, this is a romanticized picture of how *Terry* is actually used. At least in major metropolitan areas, stop-and-frisks typically are carried out pursuant to organized policies and

[168] *See id.* at 24.

programs, designed to suppress particular kinds of disorder.[169] Modifying those policies and programs to make use of evidence-based tactics would be less cumbersome than the romanticized picture of stop-and-frisk suggests.

To be clear, we are not advocating that the NYPD switch tomorrow to making CPW stop decisions by the numbers. Nor are we suggesting that police departments across the country start devising numerical formulas to guide all of their officers' stop-and-frisk decisions. As we discuss below, though, there are many midpoints between rigid, mechanical selection of all suspects to be subjected to stop-and-frisk and the current regime of subjective decision-making almost entirely unstructured by explicit criteria. We think it makes sense for scholars, police executives, and elected officials to begin considering how statistical techniques such as SHR analysis can produce evidence-based best practices to guide—perhaps just presumptively—at least some categories of *Terry* stops.

IV. Bringing *Terry* into the Information Age

As we have already suggested, SHR analysis could conceivably be used in several different ways: it could help police departments simultaneously improve both the efficiency and the fairness of their stop-and-frisk practices, it could be used by courts assessing whether a police department has engaged in illegal discrimination, and it could be used by courts assessing whether particular stops were supported by "reasonable articulable suspicion." SHR analysis could be put to any of these uses without any changes in existing legal rules, but there are

[169] *See* Meares, *supra* note 33.

changes that would help. As we discuss below, identifying the obstacles that would be encountered in trying to make use of SHR analysis to combat police discrimination is a useful way to highlight certain respects in which police practices and legal doctrines lag developments in information technology and statistical analysis.

A. Improving the Efficiency and Fairness of Stop-and-Frisk

SHR analysis could help police departments simultaneously improve both the efficiency and the fairness of their stop-and-frisk practices by reducing categories of stops that have both a relatively large racially disparate impact and a relatively low probability of "success"—i.e., a low likelihood of turning up what the police are looking for. For CPW stops, SHR analysis can help the police drastically reduce the number of stops they carry out and reduce the racially disproportionate impact of those stops without significantly reducing the number of guns they recover. It could help the police do this in either of two ways. First, the police could develop statistical heuristics (like the *Terry* score we described above) that officers could apply on the street, perhaps only presumptively, in deciding which suspects to subject to stops. Second, if the logistical difficulties of having officers apply an arithmetic heuristic are prohibitive, police managers could use SHR analysis to help identify categories of stops that are both inefficient (in the sense that they result relatively rarely in the discovery of weapons) and racially burdensome (in the sense that they impose an unusually high disparate impact on racial minorities), and then encourage their officers to avoid stops that fall within those categories. The instruction might be something like, "It looks like we're stopping lots of people based on 'furtive movements,' particularly in Brownsville, and those stops aren't recovering many weapons and tend

disproportionately to involve black or Hispanic suspects. So let's try to pull back on stops of that kind, particularly in Brownsville."

We can think of at least two reasons why the police might resist using SHR analysis even in this after-the-fact, course-correcting way. First, the police might doubt that SHR analysis accurately characterizes the likelihood a stop will result in the discovery of a weapon. Second, the real point of the stops might be for purposes other than discovering weapons: the point might be to deter people from carrying weapons, or to send a message that the police, rather than criminals, control the streets. As to the first concern, the evidence is strong that, at least for CPW stops, the SHR does a better job of identifying stops likely to be successful than the police do on their own. As to the second, perhaps it should not need saying, but an interest in general deterrence or sending a message cannot justify a *Terry* stop in the absence of particularized suspicion. Even when there is an objective basis for reasonable suspicion, a desire to demonstrate the power and authority of the police is a dubious objective for stop-and-frisk; the Supreme Court noted in *Terry* that pedestrian stops are likely to be particularly damaging to police-community relations when they are used for that purpose.[170] At a minimum, if a police

[170] 392 U.S. at 14 n.11. The test for the legality of a *Terry* stop is objective; it depends on the grounds for reasonable articulable suspicion, not on the actual motivation for the stop. *See supra* text at notes 65-69. Still, using stop-and-frisk for purposes of general deterrence, or in order to reinforce the power and authority of the police, seems contrary to the spirit of *Terry*, particularly given the Court's warning about the damaging effects of stops "motivated by the officers' perceived need to maintain the power image of the beat officer." 392 U.S. at 14 n.11 (quoting LAWRENCE P. TIFFANY, DONALD M. MCINTYRE & DONALD L. ROTENBERG, DETECTION OF CRIME: STOPPING AND QUESTIONING, SEARCH AND SEIZURE, ENCOURAGEMENT AND ENTRAPMENT 47-48 (1967).
General deterrence is different, of course, from stopping a crime already in progress or about to be commenced. The latter is a paradigmatically appropriate reason for a stop-and-frisk; it was the very rationale for the stop in *Terry* itself. The point of the stop-and-frisk power recognized in *Terry* was to allow a police officer "to take swift measures to discover the true facts" in a

department wishes to use stop-and-frisk selectively for *in terrorem* objectives, it should defend its policy publicly and explicitly.

B. Judicial Assessments of Discriminatory Intent

As we have discussed above, courts demand a showing of "total or seriously disproportionate" impact—like in *Yick Wo*—for statistical evidence to satisfy the inference of discriminatory intent. In *Floyd*, Judge Scheindlin avoided this requirement by finding independent evidence of discriminatory intent from police department statements and training materials, thus obviating the need to infer intent from statistical evidence for purposes of the Fourteenth Amendment.[171] Because SHR analysis makes use of all information that officers have before making a stop (or, at least, all such information that they enter into the department's records), it allows researchers to control for a wide range of racially neutral factors that may contribute to the stop decision.[172] When discriminatory intent exists, therefore, SHR analysis will often be able to provide strong evidence of it. Furthermore, SHR analysis avoids the benchmarking problem that has so often doomed equal protection challenges to police practices: it does not require identifying, at the outset, the racial distribution that would be expected if stops were conducted without discrimination.[173] Still, the evidence will be circumstantial, so its legal

---

particular situation and to "neutralize the threat of harm if it materialized." 392 U.S. at 30.

[171] *See supra* n. ___.

[172] In rejecting statistical proof of discrimination, the Supreme Court has often stressed the need to rule out race-neutral explanations for disparate impact, *see, e.g.*, *Davis*, 426 U.S. at 246; *Feeney*, 442 U.S. at 273-75, and Justice Brennan's dissent in *McCleskey* noted the usefulness of regression analysis for just this purpose, *see* 481 U.S. at 327.

[173] *See, e.g.*, Chavez v. Ill. State Police, 251 F.3d 612 (7th Cir. 2001); United States v. Alcaraz-Arellano, 302 F. Supp. 2d 1217 (D. Kan. 2004).

significance will depend on judicial willingness to rely on evidence of this kind in finding discriminatory intent.

As we have noted earlier, lower courts have signaled an increasing receptivity toward statistical evidence of police discrimination. And SHR analysis has three features that should appeal to courts that have been wary of relying on less powerful kinds of statistical proof. First, it can provide evidence—albeit circumstantial—of discrimination *in particular cases*. Second, it can provide a strong rebuttal to the most common "neutral" explanation for racial disparities in stop rates: more aggressive policing in "high crime" areas. Third, because SHR analysis can be used by police departments to increase the efficiency of *Terry* stops while simultaneously decreasing the racially disproportionate impact of the stops, the *failure* of a police department to make use of the lessons of SHR analysis may itself, in some instances, be evidence of discriminatory intent. We address these points in turn.

1. Proof of discrimination in particular cases. The Supreme Court in *McCleskey* stressed the defendant's failure to offer evidence "specific to his own case that would support an inference that racial considerations played a part in his sentence,"[174] and seemed concerned that efforts to develop evidence of that kind would inevitably intrude on the confidential decision-making of juries, judges, and prosecutors. Lower courts addressing claims of discriminatory policing have largely followed the lead of *McCleskey* and have been reluctant to make a finding of discrimination absent evidence that "a police officer decided to approach [the defendant]

[174] 481 U.S. at 292-93.

solely because of his or her race."[175] Even in *Floyd*, Judge Scheindlin found an individual equal protection violation only in the case of one of the nineteen named plaintiffs, Cornelio McDonald,[176] because McDonald was the only plaintiff with evidence that police targeted him instead of other, non-black individuals who were in the same place and "behaving no differently," but who were not stopped.[177]

SHR analysis provides another avenue for showing discrimination in particular cases. By pointing to a low *ex ante* probability that a stop would result in the discovery of what the police are looking for, SHR analysis can suggest that a stop was not justified by neutral factors. That will not prove conclusively, of course, that the stop was motivated by racial animus, but courts have suggested that the defendant need only make a "preponderance of the evidence" showing that such considerations motivated his stop.[178] In some cases the low probability that a stop would turn up evidence of criminal activity might suffice to meet that standard, alone or in conjunction with other evidence. Such an argument may be particularly persuasive when the stop

[175] United States v. Avery, 137 F.3d 343, 355 (1997) (internal citations omitted).
[176] 959 F. Supp. 2d at 632-33.
[177] *Id*.
[178] *Avery*, 137 F.3d at 355. Dicta in *Avery* also suggested, in accord with some other decisions, that plaintiffs claiming discrimination in violation of the Equal Protection Clause must show that the police approached them "solely because of" their race. *Id.* (quoting United States v. Travis, 62 F. 3d 170, 174 (6th Cir. 1995)). Taken literally, that would make the constitutional bar against discrimination virtually meaningless. Even the most racist police officer will almost always take at least some other factors—such as age, gender, and behavior—into account in identifying suspects. Stops motivated by race violate equal protection notwithstanding the involvement of such other factors. See*, e.g.*, Hunter v. Underwood, 471 U.S. 222, 225 (1985) (affirming lower court's ruling that, "[t]o establish a violation of the fourteenth amendment in the face of mixed motives, plaintiffs must prove by a preponderance of the evidence that racial discrimination was *a substantial or motivating factor*. . .") (emphasis added); United States v. City of Yonkers, 96 F.3d 600, 611-12 (2d Cir. 1996) ("The plaintiff need not show, however, that a government decisionmaker was motivated solely, primarily, or even predominantly by concerns that were racial. . . Rather, the plaintiff need begin only by showing that race was 'a motivating factor.'") (emphasis in original).

occurs in an area with a relatively high threshold for conducting stop-and-frisks. For example, if a black or Hispanic individual with low *ex ante* hit rate is stopped on the Upper West Side—a predominantly white neighborhood with a high bar for conducting stops—it is at least suggestive that the stop was racially motivated. And perhaps that should be enough to shift the burden to the government to disprove discriminatory intent.[179]

2. Aggressive policing in "high crime" areas. The most plausible neutral explanation for racial disparities in stop rates will often lie in differences in policing tactics across locations. Because areas with high incidence of violent crime are often also those areas with heavy use of stop-and-frisk,[180] it will often be plausible to argue that aggressive policing in these areas—which happen to be predominately minority—is justified by the legitimate goal of crime reduction. This would make the resulting disparity incidental, rather than purposeful, and therefore not unconstitutional. In fact, this is the position taken by the Department of Justice; in its most recent guidelines, the DOJ emphasized that aggressive law enforcement efforts in "high crime areas" were proper if supported by "reliable, empirical data."[181] As an example, the guidelines address a situation in which the majority of a department's drug arrests occur in heavily minority precincts: so long as the police do not act based on racial animus, "officers can properly decide to enforce all laws aggressively in that area, including less serious quality of life

[179] Burden shifting regarding discriminatory motive is, of course, familiar in other contexts, including Title VII litigation, *see* McDonnell Douglas Corp. v. Green, 411 U.S. 792, 804-05 (1973), and peremptory challenges in jury selection, *see* Batson v. Kentucky, 476 U.S. 79, 96-97 (1986).

[180] *See, e.g.*, Goel, Rao & Shroff, *supra* note 25, at 13.

[181] U.S. DEP'T OF JUSTICE, GUIDANCE FOR FEDERAL LAW ENFORCEMENT AGENCIES REGARDING THE USE OF SEX, RACE, ETHNICITY, GENDER, NATIONAL ORIGIN, RELIGION, SEXUAL ORIENTATION, OR GENDER IDENTITY (2014), *available at* <www.justice.gov/sites/default/files/ag/pages/attachments/2014/12/08/use-of-race-policy.pdf>.

ordinances, as a means of increasing drug-related arrests."[182] That suggests a police department would be acting constitutionally if it adopted a policy of conducting stops on the basis of slightly lower suspicion—though still sufficient to meet the *Terry* standard—in high crime neighborhoods.

The guidelines, however, also suggest that courts should take a searching look at the "high crime" rationale, and some appellate courts have echoed that warning. The Ninth Circuit, for example, has warned that the judiciary "must be particularly careful to ensure that a 'high crime' area factor is not used with respect to entire neighborhoods or communities in which members of minority groups regularly go about their daily business, but is limited to specific, circumscribed locations where particular crimes occur with unusual regularity."[183] The DOJ guidelines' emphasis on "reliable, empirical data" underscores this point: if police departments bring law enforcement efforts to bear more heavily upon minority communities, they must do so on the basis of specific evidence that such enforcement is necessary to reduce crime. That is not always the case. In *Floyd*, for example, the plaintiffs' liability expert concluded from a review of UF-250s that officers often used "high crime area" to describe entire boroughs, and that the incidents for which they checked the box did not correlate to actual crime rates in those areas.[184]

SHR analysis facilitates a fine-grained examination of the degree to which racial disparities in stop-and-frisk practices result from a legitimate focus on areas of high crime. In

[182] *Id*. at 5.
[183] United States v Montero-Camargo, 208 F.3d 1122, 1138 (9th Cir. 2000). *See also* United States v. Caruthers, 458 F.3d 459, 467 (6th Cir. 2006) (noting that "labeling an area 'high-crime' raises special concerns of racial, ethnic, and socioeconomic profiling.)
[184] *See supra* note 158.

New York, as we explained above, SHR analysis shows that officers were indeed stopping individuals on the basis of significantly less evidence in high crime neighborhoods, which may have been justified, at least in theory, as an appropriate effort at crime control. However, SHR analysis further indicates that *within* nearly every location, blacks and Hispanics were typically stopped on the basis of less evidence than whites. Thus, the disparate racial impact of stop-and-frisk in New York was not driven solely by more aggressive policing in high-crime areas; moreover, it is difficult to come up with a neutral, non-discriminatory explanation for the within-neighborhood disparities. And if the police department *did* have a policy to require different levels of suspicion for stops in different neighborhoods, that policy should have been explained and justified.

3. Failure to respond to SHR analysis. Big data, coupled with advanced statistical techniques such as SHR, can provide another basis for equal protection challenges to discriminatory stop-and-frisk practices. Because methods like SHR can help police departments significantly reduce the racially disproportionate impact of their practices, with few (if any) adverse consequences for the objectives of the program, the failure of police departments to take advantage of these opportunities can and should be seen as rebuttal proof of an equal protection violation.

Perhaps a police department has no duty to carry out SHR analysis on its own. But if third parties carry out the analysis and demonstrate to the police that the disparate impact of stops could be decreased and the efficiency of stops increased by avoiding certain categories of stops with low SHRs, and the police response is to carry on business as usual, the failure to

respond might be seen as a violation of equal protection. If, for example, the core of equal protection is a right to "equal concern and respect"—as many scholars have argued[185]—there is a straightforward argument that this right is denied when a police department does not bother to take steps that would make its practices less disproportionately burdensome to members of traditionally disadvantaged minorities and simultaneously more effective at achieving their stated objectives.

The Supreme Court, and most lower courts, have so far refused to recognize any liability for the failure to make adjustments in response to disparate impact outside the context of Title VII. In fact the Supreme Court has said explicitly that "'discriminatory purpose' implies more than intent as awareness of consequences."[186] Lower courts have cited this language in rejecting arguments that the selection of a particular community for an enforcement operation constituted discrimination just because it was foreseeable that minority group members would be disproportionately impacted.[187] Nonetheless some trial courts in recent years have suggested

[185] RONALD DWORKIN, TAKING RIGHTS SERIOUSLY 272-23 (1977); *see also, e.g.*, Matthew D. Adler, *Rights Against Rules: The Moral Structure of American Constitutional Law*, 97 MICH. L. REV. 1, 116 & n. 386 (1998); Ronald Dworkin, *Unenumerated Rights: Whether and How Roe Should Be Overruled*, 59 U. CHI. L. REV. 381, 387-88 (1992); Kim Forde-Mazrui, *Tradition as Justification: The Case of Opposite-Sex Marriage*, 78 U. CHI. L. REV. 281 (2011).

[186] Wayte v. United States, 470 U.S. 598, 610 (1985) (ellipses omitted) (quoting Personnel Administrator v. Feeney, 442 U.S. 256, 279 (1979)).

[187] United States v. Turner, 104 F.3d 1180, 1184-85 (9th Cir. 1997); see also United States v. Thorpe, 2007 WL 4239201, at *1 (E.D. Mich. Dec. 3, 2007) (finding that plaintiff could not satisfy discriminatory intent requirement in selective prosecution claim because "the caselaw does not support [the] argument" that "the government's pursuit of a program despite knowledge of that program's discriminatory effect is by itself 'some evidence' of discriminatory intent"); Pollard v. Wawa Food Market, 366 F. Supp. 2d 247, 253 (E.D. Pa. 2005) (dismissing plaintiffs' section 1981 claim on the ground that it did not allege intentional discrimination where allegations were based on the fact that defendant intentionally applied policy that it knew would have disparate impact, because "discrimination is intentional only if the defendant selected or

greater openness to entertaining equal protection claims based on the failure to take corrective measures to address known disparities. For example, the District of New Jersey in 2003 upheld a plaintiff's complaint of intentional discrimination where a federal agency "was well aware of the potential disproportionate and discriminatory burden placed upon that community and failed to take measures to assuage that burden."[188]

Moreover, Judge Scheindlin's opinion in *Floyd* is itself now precedent for extending "deliberate indifference" liability to cases of disparate impact by framing them as instances of failure to properly train and supervise government employees.[189] This kind of liability traditionally has been based upon a municipality's failure to train its employees with regard to specific policies that caused constitutional violations[190]—and because disparate impact by itself has generally been insufficient to make out a constitutional violation, such liability has rarely been extended to failures to correct disparities alone. But *Floyd* may signal a growing

---

reaffirmed a particular course of action 'because of,' not merely 'in spite of,' its adverse effects upon an identifiable group").

[188] South Camden Citizens in Action v. N.J. Dep't of Envtl. Prot., 254 F. Supp. 2d 486, 495 (D.N.J. 2003); *see also* Garvey ex rel. Doe v. Unified Sch. Dist. *262*, 2005 WL 2548332, at *2 (D. Kan. Oct. 12, 2005) (denying defendant-district's motion to dismiss—"albeit reluctantly"—where plaintiff's claim of intentional discrimination was based on the argument that "the district was notified [of the policy's adverse impact] and . . . did nothing to resolve the impact the policy had on minorities"); Almendares v. Palmer, 284 F. Supp. 2d 799, 808 (N.D. Ohio 2003) (denying motion to dismiss where plaintiffs alleged that defendants purposely discriminated against them by choosing to continue policy despite knowledge that the policy harmed minority food stamp applicants, because "one could logically infer" from these facts that policy was being implemented "because of" its impact on national origin).

[189] 959 F.Supp.2d at 564, 590.

[190] *See, e.g.*, City of Canton v. Harris, 489 U.S. 378, 388-89 (1989) ("[A] municipality can be liable under § 1983 only where its policies are the moving force [behind] the constitutional violation. Only where a municipality's failure to train its employees in a relevant respect evidences a "deliberate indifference" to the rights of its inhabitants can such a shortcoming be properly thought of as a city 'policy or custom' that is actionable under § 1983.") (internal quotation marks omitted).

recognition that, in an age of big data, allowing unjustified racial disparities to persist, unchallenged, can itself constitute invidious discrimination.

C. Judicial Assessments of Fourth Amendment Reasonableness.

Beyond its implications for equal protection analysis, SHR analysis and other, similar uses of big data could and should inform judicial applications of the Fourth Amendment guarantee against "unreasonable searches and seizures." Because SHR analysis can estimate the *ex ante* probability that any particular *Terry* stop would wind up finding a weapon—or any other contraband or evidence of criminal activity—it raises the possibility of an objective, quantitative test for "reasonable articulable suspicion." The Supreme Court has resolutely refused in the past to attach any numerical probability threshold to "reasonable articulable suspicion" or to reduce it to any rigid formula; it has pegged the standard instead, to "commonsense judgments and inferences about human behavior."[191] In doing so, though, the Court has stressed the unavailability of "empirical studies dealing with inferences drawn from suspicious behavior."[192] Big data, combined with techniques such as SHR analysis, is beginning to fill that gap.

Moreover, there is a natural affinity between the SHR technique and the requirement of "reasonable, articulable suspicion." Faced with a claim that a particular stop was unreasonable because the SHR was too low, law enforcement can be expected to respond that the regression model through which the SHR was calculated does not capture everything the officer relied upon in deciding to conduct a stop-and-frisk. But the SHR method relies on the information that the

[191] *Illinois v. Wardlow*, 528 U.S. 119, 124-25 (2000).
[192] *Id.*

police department itself chooses to record—the information that, in effect, it has the officer "articulate" at the time of the stop. So if that information is insufficient to capture the basis for the stop, the police department can and should revise its form. Perhaps, for example, recording "furtive movements" as a binary variable—either there are furtive movements or there are not—fails to capture the fact that some furtive movements are more furtive, and hence more probative, than others. But then a form that fails to distinguish *really* furtive movements from *kind-of* furtive movements is failing to have officers articulate the full basis for the stop.

Nonetheless there is an obvious difficulty in trying to quantify reasonable suspicion. It is the problem of the threshold: how low a probability is too low to satisfy the Fourth Amendment?[193] The Supreme Court has said that reasonable suspicion can be based on facts that fall "considerably short of satisfying a preponderance of the evidence standard,"[194] but that tells us only that the threshold falls appreciably below 50%. There is, however, precedent for defining reasonableness in terms of the probability that a search will uncover evidence of illegality. The strongest precedent comes not in the context of stop-and-frisk but in the context of searches and seizures in furtherance of "special needs" other than general crime control—in particular, in the

[193] In addition, there is the risk that once a threshold is selected it will be applied too mechanically—that, for example, judges will fail to consider the possibility that evidence not reported by the police made the probability of finding contraband far lower than the record suggests. For reasons of this kind, Orin Kerr has argued against quantifying probable cause: he worries that doing so would make judges reviewing warrant applications less inclined to speculate about evidence the police left out. Orin Kerr, *Why Courts Should Not Quantify Probable Cause*, *in* THE POLITICAL HEART OF CRIMINAL JUSTICE: ESSAYS ON THEMES OF WILLIAM J. STUNTZ 131 (Michael Klarman et al. eds, 2012). Whatever may be the case with regard to warrants, though, it is far from clear that courts reviewing *Terry* stops do much speculating of this kind today, or that it would be helpful for them to do so. Particularly given the programmatic nature of stop-and-frisk, we think more is to be gained by increasing the rigor with which reasonable articulable suspicion is assessed.

[194] *United States v. Arvizu*, 534 U.S. 266, 274 (2002)

context of highway checkpoints. Like other "special needs" searches and seizures, the constitutionality of sobriety stops and immigration checkpoints depends on the reasonableness of the *program* under which they are implemented, and the reasonableness of the program depends in part on the hit rate. The Supreme Court has upheld a sobriety checkpoint with a 1.6% hit rate, reasoning that detecting drunk drivers at this frequency could "reasonably be said" to advance the state's legitimate interests in highway safety,[195] and it earlier upheld an immigration checkpoint with a 0.1% hit rate.[196]

However, in both cases, the Court emphasized the minimal nature of the intrusion resulting from routine checkpoint stops, where "checkpoints are selected pursuant to the guidelines, and uniformed police officers stop every approaching vehicle."[197] In the street encounter context, where stops involve a greater degree of intrusion and possibilities for officer abuse,[198] such numbers may be—and we think they probably are—too low to outweigh the liberty and privacy interests at stake. Justice Scalia once suggested that a single traffic violation could give rise to a reasonable suspicion of drunk driving only if the chances were at least "1 in 20" that the driver was intoxicated—a hit rate of 5%.[199] And, in *Anderson v. Cornejo*, the Seventh Circuit airport search case discussed earlier, Judge Easterbrook of the Seventh Circuit suggested that a 27.6% hit rate was "not to be sneezed at,"[200] but that suggests that a hit rate of

---

[195] *Michigan Dep't of State Police v. Sitz*, 496 U.S. 444, 455 (1990).
[196] *United States v. Martinez-Fuerte*, 428 U.S. 543, 554 (1976) (noting that deportable aliens were found in 171 of 146,000 vehicles that passed through a checkpoint).
[197] *Id*. at 453; *Martinez-Fuerte*, 428 U.S. at 558.
[198] *See Martinez-Fuerte*, 428 U.S. at 559 (suggesting that stops made by officers outside of checkpoints provide more room for "abusive and harassing stops").
[199] Navarette v. California, 134 S. Ct. 1683, 1695 (2014) (Scalia, J., dissenting).
[200] 355 F.3d at 1021. Judge Easterbrook asked rhetorically, "[i]f about 0.1% of black women returning from foreign travel are smuggling, and the agents select so carefully that 28% of those

1% or even in the range of 3%-4%—the range of the SHRs for at least some of the *Floyd* plaintiffs—might not be high enough to make pedestrian stops "reasonable" within the meaning of the Fourth Amendment.

The broader point is that the level of suspicion required to make investigative stops "reasonable" depends in part on the importance of the goals the government is pursuing, in part on how intrusive and burdensome the stops are, and in part on whether there are other, less objectionable ways to further the government's goals. These are the factors the Supreme Court has routinely considered—along with hit rates—in assessing the constitutionality of highway checkpoints under the "special needs" doctrine. For example, in *Delaware v. Prouse*, when the Supreme Court struck down a program calling for stopping cars randomly on the highway to check for licenses, the Justices explained that "[t]he marginal contribution to roadway safety possibly resulting from a system of spot checks cannot justify subjecting every occupant of every vehicle on the roads to a seizure—limited in magnitude compared to other intrusions but nonetheless constitutionally cognizable."[201] So, too, the costs and benefits of a stop-and-frisk program would need to be considered in determining a numerical threshold for reasonable suspicion. Once again, though, the key move would be to view stop-and-frisk programs *as programs*, and to assess their reasonableness at the programmatic level, just as has long been done for "special needs" searches.

---

searched are caught with contraband, where's the beef?" That suggests that the reasonableness of a stop-and-frisk program might be judged in part by comparing the hit rate with estimates of how common the pertinent criminal behavior is in the population from which suspects are selected.

[201] 440 U.S. 648, 661 (1979).

A further point: once stop-and-frisk is assessed as a program, it becomes especially difficult to disregard a disproportionate impact on minority communities in counting its costs. When the Supreme Court ruled in *Whren* that discriminatory intent is irrelevant to the reasonableness of an arrest under the Fourth Amendment, it relied heavily on the idea that Fourth Amendment analysis typically does not inquire into "the actual motivations of individual officers."[202] *Whren* has been heavily criticized,[203] but it had at least this to recommend it: inquiries into motive are often messy and inconclusive. Objective tests are easier to apply. And there is no objective test for the racial fairness of an individual search or seizure.

Programs are different. It doesn't make sense to ask whether an individual *Terry* stop has a disparate impact; it *does* make sense to ask that about a program, and disparate impact is a completely objective test. Nor is disparate impact important only as evidence of the subjective motivation—as evidence, that is, of discriminatory intent. Thoughtful and fair-minded criminal justice officials seek, when possible, to avoid policies that disproportionately burden racial minorities even when those policies stem from nondiscriminatory motives. They do so because disproportionately burdening traditionally disadvantaged groups is a bad thing to do: it piles burdens on the already burdened, which itself seems worth avoiding, it poisons relations between law enforcement and minority communities, and it corrodes the perceived legitimacy of the law. None of this means that any criminal justice policy that disproportionately burdens racial minorities should be abandoned; sometimes disparate impact is a necessary evil. It does mean, though, that in assessing the "reasonableness" under the Fourth Amendment of a *program* of searches or seizures, it should weigh against a finding of reasonableness that the program

[202] 517 U.S. at 813.
[203] Including by one of us. *See* Sklansky, *supra* note 5.

disproportionately burdens racial minorities, or any other traditionally disadvantaged group. Disparate impact of this kind is a cost, and it should be taken into account in assessing reasonableness. Big data and new statistical tools like SHR analysis make it increasingly feasible to do this, and they make *not* doing it increasingly indefensible.

CONCLUSION

Big data is changing policing. It is providing law enforcement agencies with new tools that carry with them new risks. But big data also provides opportunities to create new forms of police accountability: new ways to monitor and assess how the police do their work, and to help them to improve the fairness and effectiveness of their tactics. We have focused here on the possibilities of this kind provided by combining stop-and-frisk data with a particular, new statistical tool—SHR analysis—but we have also tried to make three broader points. First, the potential uses of big data to make policing both fairer and more effective are just beginning to be discovered. Second, these new tools of police accountability warrant reexamination of some traditional rules and assumptions pertaining to legal oversight of the police. Third and finally, *Terry* stops—both pedestrian pat-downs and investigatory traffic stops—should be analyzed as policies, not as isolated interactions. That is how they usually are implemented, that is how they tend to be experienced, and that is how to make best use of the tools of oversight provided by big data.