2023

Are Police Officers Bayesians? Police Updating in Investigative Stops

Jeffrey A. Fagan
Columbia Law School, jfagan@law.columbia.edu

Lila J.E. Nojima
Columbia Law School

Recommended Citation

Jeffrey A. Fagan & Lila J. Nojima, *Are Police Officers Bayesians? Police Updating in Investigative Stops*, 113 J. Crim. L. & Criminology 593 (2023).
Available at: https://scholarship.law.columbia.edu/faculty_scholarship/3970

0091-4169/23/11303-0593
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY Vol. 113, No. 3

CRIMINOLOGY

ARE POLICE OFFICERS BAYESIANS? POLICE UPDATING IN INVESTIGATIVE STOPS

JEFFREY FAGAN[*] & LILA J.E. NOJIMA[**]

Theories of rational behavior assume that actors make decisions where the benefits of their acts exceed their costs or losses. If those expected costs and benefits change over time, the behavior will change accordingly as actors learn and internalize the parameters of success and failure. In the context of proactive policing, police stops that achieve any of several goals—constitutional compliance, stops that lead to "good" arrests or summonses, stops that lead to seizures of weapons, drugs, or other contraband, or stops that produce good will and citizen cooperation—should signal to officers the features of a stop that increase its rewards or benefits. Having formed a subjective estimate of success (i.e., prior beliefs), officers should observe their outcomes in subsequent encounters and form updated probability estimates, with specific features of the event, with a positive weight on those features. Officers should also learn the features of unproductive stops and adjust accordingly. A rational actor would pursue "good" or "productive" stops and avoid "unproductive" stops by updating their knowledge of these features through experience.

[*] Isidor and Seville Sulzbacher Professor of Law, and Professor of Epidemiology, Columbia University. Thanks to Greg DeAngelo, Chris Wildeman, Amanda Geller, John MacDonald, participants at CELS 2019, and anonymous reviewers for very helpful comments. Jeffrey Fagan served as expert witness for plaintiffs in Floyd v. City of New York, 959 F. Supp. 2d 540 (S.D.N.Y. 2013); Ligon v. City of New York, 959 F.3d 166 (2d Cir. 2013); and Davis v. City of New York, 902 F. Supp. 2d 405 (S.D.N.Y. 2012). Fagan's research for this Article was made possible in part by The Madsen Family Faculty Research Fund at Columbia Law School. All opinions and errors are solely ours. Correspondence to: Jeffrey Fagan, Columbia Law School, 435 W. 116th Street. New York, NY 10027. Tel. 212 854 2624, jfagan@law.columbia.edu.

[**] J.D. 2020, Columbia Law School.

We analyze data on 4.9 million Terry *stops in New York City from 2004–2016 to estimate the extent of updating by officers in the New York Police Department. We compare models using a frequentist analysis of officer behavior with a Bayesian analysis where subsequent events are weighted by the signals from prior events. By comparing productive and unproductive stops, the analysis estimates the weights or values—an experience effect—that officers assign to the signals of each type of stop outcome. We find evidence of updating using both analytic methods, although the "hit rates"—our measure of stop productivity including recovery of firearms or arrests for criminal behavior—remain low. Updating is independent of total officer stop activity each month, suggesting that learning may be selective and specific to certain stop features. However, hit rates decline as officer stop activity increases. Both updating and hit rates improved as stop rates declined following a series of internal memoranda and trial orders beginning in May 2012. There is also evidence of differential updating by officers conditional on a variety of features of prior and current stops, including suspect race and stop legality. Though our analysis is limited to NYPD stops, given the ubiquity of policing regimes of intensive stop and frisk encounters across the United States, the relevance of these findings reaches beyond New York City. These regimes reveal tensions between the* Terry *jurisprudence of reasonable suspicion and evidence on contemporary police practices across the country.*

INTRODUCTION

In *Terry v. Ohio*, the Supreme Court carved out an exception to the probable cause requirement of the Fourth Amendment—officers were permitted to stop an individual based on a lower showing of reasonable suspicion.[1] The Court described an officer who "in light of his experience" could "reasonably conclude . . . that criminal activity may be afoot and that the persons with whom he is dealing may be armed and presently dangerous . . . ."[2] The Court, however, emphasized that such reasonable suspicion must not be based on an officer's "inchoate and unparticularized suspicion or 'hunch,' but [on] the specific reasonable inferences which he is entitled to draw from the facts in light of his experience."[3] Experience, not mere hunches, should guide officers in their discretion to stop and frisk individuals. But what exactly does this experience entail and does it meaningfully differ from a mere hunch? This paper uses novel officer-level data to assess the extent of officer learning and updating over time. We investigate whether an officer's past successes predict her probability of success in subsequent stops, and whether officers improve their stop accuracy over time.[4] In sum, does experience help reduce the likelihood of unconstitutional and unreasonable stops, as the *Terry* Court posited?

Theories of rational behavior assume that actors will maximize their returns by making decisions where the benefits of their acts exceed their costs or losses.[5] If those expected costs and benefits change over time, the behavior will change accordingly as actors learn and internalize the parameters of success and failure. We apply this perspective to examine a core feature of the "new policing": the stop "careers" of officers working in regimes of

[1] 392 U.S. 1, 31 (1968).

[2] *Id.* at 30–31.

[3] *Id.* at 27–28.

[4] *See* Max Minzner, *Putting Probability Back into Probable Cause*, 87 TEX. L. REV. 913, 930 (2009).

[5] *See generally* GARY S. BECKER, THE ECONOMIC APPROACH TO HUMAN BEHAVIOR 8–14 (1976) (stating a theory of human behavior based on rational weighing of alternatives within an entity's utility function).

intensive investigative stops, or "stop and frisk" encounters with civilians.[6] Here, we investigate these questions using both publicly available data and data obtained from the New York Police Department (NYPD).[7] This policing tactic, however, is not unique to the NYPD. It has been adopted by police departments across the U.S., generated extensive empirical literature, and been the focus of civil rights litigation resulting in court oversight in several places.[8]

In practice, investigative stops that achieve any of several goals—constitutional compliance, stops that lead to "good" arrests or summonses, stops that lead to seizures of weapons, drugs, or other contraband (stolen property), or stops that produce good will and citizen cooperation—should signal to officers the features of a stop that increase those rewards or benefits. In theory, officers, having formed a subjective estimate of success (i.e., prior beliefs) through prior activity periods, will observe their outcomes and update their probability estimates, incorporating specific features of the event and placing a positive weight on those features. Officers should also learn the features of unproductive stops and adjust their decision making accordingly. A rational actor would pursue productive or good stops where firearms are seized or offenders are arrested, and avoid unproductive stops that yield neither, by updating their knowledge of these features through experience. A positive updating of stop activity through learning should increase the social good of policing by improving public safety and calling offenders to account.

This Article extends the empirical literature on stop and frisk by analyzing the stop "careers" of officers to assess the extent of learning and updating over time. We construct a database on the outcomes of officer stops over discrete but consecutive incidents across an officer's stop "career." We

[6] Philip B. Heymann, *The New Policing*, 28 FORDHAM URB. L.J. 407, 422–24 (2000); *see also* NAT'L RES. COUNCIL, PROACTIVE POLICING: EFFECTS ON CRIME AND COMMUNITIES 29–33 (2018) (reviewing the evolution of a saturated and aggressive policing model widely common to many American police departments); Rachel A. Harmon, *The Problem of Policing*, 110 MICH. L. REV. 761, 776–81 (2012) (arguing for analyses and scholarship that show constitutional law can regulate police to maximize the returns from policing.).

[7] These data contain records for each stop and frisk event made by the NYPD during the study period, January 2004 to December 2016, as recorded on UF-250 forms. Data obtained through discovery in the *Floyd v. City of New York*, 959 F. Supp. 2d 540 (S.D.N.Y. 2013), litigation included an encrypted officer identifier for each stop. *See infra* Section II.A.

[8] State v. Soto, 734 A.2d 350, 360–61 (N.J. Super. Ct. Law Div. 1996); *e.g.*, Settlement Agreement, Class Certification, and Consent Decree, Bailey v. City of Philadelphia, No. 10-v-5952-SD (E.D. Pa. June 21, 2011); Consent Decree, United States v. City of Los Angeles, No. 00-cv-11769(GAF)(RC) (C.D. Cal. Jun. 15, 2001); Investigatory Stop and Protective Pat Down Settlement Agreement (2015), https://www.aclu-il.org/sites/default/files/wp-content/uploads/2015/08/2015-08-06-Investigatory-Stop-and-Protective-Pat-Down-Settlement-Agreeme . . . .pdf [https://perma.cc/RB43-DGHM].

extend the “hit rate”[9] literature by examining whether success rates of officers vary over time, whether there is evidence of learning that will lead to more productive and constitutionally sound investigative stops of citizens, and whether there are dimensions of learning that are specific to reasonable suspicion stops of different categories of persons.

Prior hit rate tests often rely on single incidents or stops, or aggregates of those stops within officers, and compare their outcomes—frisks, searches, citations, arrests, contraband seizures, weapons seizures—across individuals or places.[10] Each stop is an event in this design, even though stops are nested within officers. Several studies have decomposed hit rates by officer race or officer-suspect race dyads. Outcomes are then compared for population groups by age, gender, or race, and on occasion, on the behavioral suspicion that formed the basis of the stop. There is valuable evidence in these studies of the behaviors of officers within a department and the preferences of a police agency in how it regards the success rates or failures of its collective officers. But studies of officers’ own decisions over time are rare and tend to

[9] Ian Ayres, *Outcome Tests of Racial Disparities in Police Practice*, 4 JUST. RES. & POL. 131, 132–35 (2002); Jeffrey Fagan, *Law, Social Science, and Racial Profiling*, 4. JUST. RES. & POL. 103, 119 (2002); John Knowles, Nicola Persico & Petra Todd, *Racial Bias in Motor Vehicle Searches: Theory and Evidence*, 109 J. POL. ECON. 203, 205 (2001) (proposing a model for assessing discrimination in suspect searches based on success rates of searches between suspects of different races); Camelia Simoiu, Sam Corbett-Davies & Sharad Goel, *The Problem of Infra-Marginality in Outcome Tests for Discrimination*, 11 ANNALS OF APPLIED STAT. 1193, 1195 (2017).

[10] See Kate Antonovics & Brian Knight, *A New Look at Racial Profiling: Evidence from the Boston Police Department*, 91 REV. ECON. & STAT. 163, 169–71 (2009); Decio Coviello & Nicola Persico, *An Economic Analysis of Black-White Disparities in the New York Police Department’s Stop and Frisk Program*, 44 J. LEGAL STUD. 315, 324–26 (2015); Ronald G. Fryer, Jr., *An Empirical Analysis of Racial Differences in Police Use of Force*, 127 J. POL. ECON. 1210, 1217–18 (2019); Jonathan Mummolo, *Modern Police Tactics, Police-Citizen Interactions, and the Prospects for Reform*, 80 J. POL. 1, 5 (2018); John A. Shjarback, David C. Pyrooz, Scott E. Wolfe & Scott H. Decker, *De-policing and Crime in the Wake of Ferguson: Racialized Changes in the Quantity and Quality of Policing Among Missouri Police Departments*, 50 J. CRIM. JUST. 42, 45–46 (2017); Simoiu et al., *supra* note 9, at 1202–03; Ravi Shroff, *Statistical Tests to Audit Investigative Stops*, 15 OHIO ST. J. CRIM. L. 565, 566–69 (2018) [hereinafter Shroff, *Statistical Tests*] (discussing hit rate tests); Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff & Sharad Goel, *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE HUM. BEHAV. 736, 737 (2020); Emma Pierson, Sam Corbett-Davies & Sharad Goel, *Fast Threshold Tests for Detecting Discrimination* 5 (Mar. 10, 2018) (unpublished manuscript), https://arxiv.org/pdf/1702.08536.pdf [https://perma.cc/PA3Q-3AFV].

aggregate officers' stop or patrol activity for between-officer comparisons.[11] It is even rarer that officers' stop activity is analyzed in succession to determine whether officers are updating their decision processes to maximize their hit rates. This Article takes up that challenge, using a within-officer design to test the accuracy and fairness of officers' decisions in succession over time.

Relying on previously unavailable unique police officer IDs,[12] we assess evidence of learning and improvement in the accuracy of NYPD officers' reasonable suspicion stops. We analyze data on 4.9 million *Terry* stops in New York City from 2004–2016 to estimate the extent of learning and updating by officers. We compare models using a linear or frequentist analysis of officer behavior with a Bayesian analysis—where subsequent events are weighted by the signals from prior events. By comparing productive and unproductive stops, the analysis estimates the weights or values—an experience effect—that officers assign to the signals of each type of stop outcome. New York is an important research site for this inquiry: the practice of *Terry* stops has been in effect for nearly three decades,[13] and there now is extensive and granular data on police stops. We use a lengthy window from 2004 to 2016, an interval during which close external scrutiny and federal civil rights litigation placed officers under close monitoring and—assuming they are rational actors—should have incentivized officers to improve their accuracy in the conduct of stops.

We find evidence of differential updating conditional on a variety of features of prior and current stops, though effect sizes remain low overall. While we find that an officer's prior month hit rates are significant positive predictors of subsequent hit rates, as are prior months' search rates, we do not find that increased stop activity in prior months is associated with increased hit rates in subsequent months. Rather, increased stop activity is associated with lower hit rates, suggesting that learning may be selective to

[11] For exceptions, see Sharad Goel, Maya Perelman, Ravi Shroff & David Alan Sklansky, *Combatting Police Discrimination in the Age of Big Data*, 20 NEW CRIM. L. REV. 181, 211–212 (2017); Greg Ridgeway & John M. MacDonald, *Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops*, 104 J. AM. STAT. ASS'N 661, 662 (2009).

[12] Stop records that included a unique encrypted officer identifier for each stop were obtained during discovery as part of the litigation in *Floyd v. City of New York*, 959 F. Supp. 2d 540 (S.D.N.Y. 2013). These data elements are not available in the publicly available stop data.

[13] *See* ELIOT SPITZER, OFF. OF THE ATT'Y GEN. OF THE STATE OF N.Y., THE NEW YORK CITY POLICE DEPARTMENT'S "STOP & FRISK" PRACTICES 88 (1999); Heymann, *supra* note 6, at 429–32; Debra Livingston, *Police Discretion and the Quality of Life in Public Spaces: Courts, Communities, and the New Policing*, 97 COLUM. L. REV. 551, 583 n.162 (1997).

specific stop features, including location and suspect characteristics, like race or ethnicity, and that there may be a ceiling on the productivity of the stop regime. Similarly, prior month frisk rates are significant positive predictors of subsequent hit rates only for weapons and contraband, but not for arrests and summons. Hit rates improved as stop rates declined following civil rights litigation that led to the first of several court orders[14] as well as internal NYPD memoranda altering stop procedures[15] beginning in March 2013.

However, the effects of the *Floyd* litigation and order on updating are unclear: we have partial evidence that stops after class certification in 2012 have a small positive association with hit rates for weapons and contraband, but that they have a negative association with hit rates for arrests and summons. These results do not lead us to the conclusion that *Terry* is bad law or that the reasonable suspicion standard should be reconsidered. Instead, they indicate a disconnect between Fourth Amendment jurisprudence and police practice, and a failure of constitutional regulation of police practices; officers may not be rational actors who consistently evaluate prior stops and learn from their experiences. While there may be some learning going on from prior stop encounters, relying on officer experience alone to guide Fourth Amendment compliance—reasonable suspicion—is insufficient to determine whether an officer had individualized and articulable suspicion. Determining constitutional compliance requires courts to conduct a more thorough and searching inquiry of each stop and give less weight to general assertions of experience or expertise.[16] Moreover, the results indicate that the institutional design of modern police departments—including quotas and institutional pressures—may undermine learning and rational behavior.

[14] *See* Floyd v. City of New York, 283 F.R.D. 153, 160 (S.D.N.Y. 2012) (concluding that plaintiffs satisfy the legal standard for class certification). Though the reasons for these changes in the ensuing months leading up to the final 2013 court order are opaque, they likely reflect a confluence of circumstances—some intended and others unintended. For instance, following the initiation of the *Floyd* litigation, there was evidence that "some officers [were] making stop without appropriately documenting them," "officers on the street may be declining to stop, question and frisk when it would be lawful [because] officers are not confident . . . [about] what they are authorized to do under the law." PETER L. ZIMROTH, FIRST REPORT OF THE INDEPENDENT MONITOR 17–18 (2015), https://www.nypdmonitor.org/wp-content/uploads/2022/09/01-MonitorsFirstReport-AsFiledInFloydDocket.pdf [https://perma.cc/ZXA3-BBVX].

[15] *See* Memorandum from James Hall, Chief of Patrol, NYPD, to Commanding Officer in Patrol Boroughs Requiring Activity Log Entries Regarding UF-250s (Mar. 5, 2013) (discussing standardization and elaboration of data elements in "Stop, Question and Frisk Report" logs).

[16] *See generally* Anna Lvovsky, *Rethinking Police Expertise*, 131 YALE L.J. 475, 497–534 (2021) (questioning the deference of courts to police experience in determining constitutional compliance).

I. BACKGROUND

A. FOURTH AMENDMENT BASICS

The Fourth Amendment provides that:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.[17]

The Amendment serves two essential functions: privacy protection and the regulation of the state.[18] The Framers, acknowledging unreasonable searches and seizures in England and the American colonies, "established the principle which was enacted into the fundamental law in the Fourth Amendment, that a man's house was his castle and not to be invaded by any general authority to search and seize his goods and papers."[19] A constitutional search and seizure, therefore, required "probable cause as a minimum requirement . . . [and] has also required the judgment of a magistrate on the probable-cause issue and the issuance of a warrant before a search is made."[20] Investigative stops, which derive their authority from the Common Law Right of Inquiry, form the legal authority for the initial contact between citizens and police.[21]

The Supreme Court's exclusionary rule gives teeth to the Fourth Amendment, "its basic functioning is clear and undisputed: evidence obtained as the result of an unconstitutional search or seizure is suppressed at trial."[22] The rule was applied to both the federal government[23] and the

[17] U.S. CONST. amend. IV.

[18] *See* Weeks v. United States, 232 U.S. 383, 390–92 (1914).

[19] *Id.* at 389–92 (describing the history of the Fourth Amendment as rooted in English law).

[20] Chambers v. Maroney, 399 U.S. 42, 51 (1970); *see also* United States v. Timms, No. 17-CR-130 (KBF), 2017 WL 3503373, at *3 (S.D.N.Y. Aug. 16, 2017) ("In most instances, the touchstone of a constitutional search is one conducted pursuant to a judicially authorized warrant, or resulting from probable cause.").

[21] *See* People v. De Bour, 352 N.E.2d 562, 571–72 (N.Y. 1976) ("[T]he common-law right to inquire, is activated by a founded suspicion that criminal activity is afoot and permits a somewhat greater intrusion in that a policeman is entitled to interfere with a citizen to the extent necessary to gain explanatory information, but short of a forcible seizure." (citations omitted)).

[22] Eugene R. Milhizer, *Debunking Five Great Myths About the Fourth Amendment Exclusionary Rule*, 211 MIL. L. REV. 211, 212 (2012).

[23] *Weeks*, 232 U.S. at 398.

states.[24] *Terry v. Ohio*, however, ushered in a sea change to Fourth Amendment jurisprudence and on-the-ground police work.[25]

1. *Terry's Escape from Probable Cause*

The facts of *Terry* are well known and summarized only in brief here.[26] In 1963, veteran officer Detective McFadden was patrolling downtown Cleveland when two men, John Terry and Richard Chilton, arose his suspicions.[27] McFadden described that based on his experience, "he had developed routine habits of observation . . . [and] 'in this case when [he] looked over they didn't look right . . . .'"[28] McFadden, suspicious that the men were "casing," approached the men, identified himself, and asked for their names.[29] They failed to answer, so he "grabbed . . . Terry, spun him around . . . and patted down the outside of his clothing."[30] McFadden felt a gun, removed Terry's coat, and retrieved the gun.[31] He subsequently patted down Chilton and a third man and discovered a gun on Chilton as well.[32] Terry and Chilton were eventually charged with carrying concealed weapons.[33]

The Supreme Court addressed the question of "whether it is always unreasonable for a policeman to seize a person and subject him to a limited search for weapons unless there is probable cause for an arrest."[34] Did the stop, frisk, and search violate Terry's Fourth Amendment rights? In the course of the opinion, the Court recognized and gave weight to the "diversity" of street encounters between police officers and the public, noting that "[t]hey range from wholly friendly exchanges of pleasantries or mutually

[24] Mapp v. Ohio, 367 U.S. 643, 654–55 (1961) (overruling Wolf v. Colorado, 338 U.S. 25 (1949)).

[25] *See* Jeffrey Fagan, Terry's *Original Sin*, 2016 U. CHI. LEGAL F. 43, 50–56 (critiquing the dilution of Fourth Amendment search thresholds following *Terry v. Ohio*, 392 U.S. 1, 31 (1968)).

[26] *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, 163 (2015); Anthony C. Thompson, *Stopping the Usual Suspects: Race and the Fourth Amendment*, 74 N.Y.U. L. REV. 956, 962–64 (1999).

[27] *Terry*, 392 U.S. at 4–7.

[28] *Id.*

[29] *Id.*

[30] *Id.*

[31] *Id.*

[32] *Id.*

[33] *Id.*

[34] *Id.* at 15.

useful information to hostile confrontations . . . ."[35] Indeed, the Court clarified that in this case, "we deal here with an entire rubric of police conduct—necessarily swift action predicated upon the on-the-spot observations of the officer on the beat—which historically has not been, and as a practical matter could not be, subjected to the warrant procedure."[36]

First, the Court provided that police stops and frisks fall within the meaning of the Fourth Amendment's searches and seizures, even if the individual is never arrested, and are therefore subject to a reasonableness inquiry.[37] Second, the Court determined that the gun seized was properly admitted as evidence against Terry.[38] Thus, the *Terry* stop was established: "carv[ing] out an exception to the 'probable cause' requirement."[39] Short of a showing of probable cause, officers are permitted to stop an individual based on an articulation of reasonable suspicion, which itself can rely on the officer's own experience.[40] Moreover, officers are entitled to conduct a frisk, or "limited search of outer clothing," when the officer reasonably believes the individual is armed and dangerous.[41] While officers are not permitted to make stops solely on "a mere 'hunch,'" the "likelihood of criminal activity need not rise to the level required for probable cause, and it falls considerably short of satisfying a preponderance of the evidence standard."[42]

[35] *Id.* at 13.

[36] *Id.* at 20.

[37] *Id.*

[38] The Court held:

> [W]here a police officer observes unusual conduct which leads him reasonably to conclude in light of his experience that criminal activity may be afoot and that the persons with whom he is dealing may be armed and presently dangerous, where in the course of investigating this behavior he identifies himself as a policeman and makes reasonable inquiries, and where nothing in the initial stages of the encounter serves to dispel his reasonable fear for his own or others' safety, he is entitled for the protection of himself and others in the area to conduct a carefully limited search of the outer clothing of such persons in an attempt to discover weapons which might be used to assault him. Such a search is a reasonable search under the Fourth Amendment, and any weapons seized may properly be introduced in evidence against the person from whom they were taken.

Id. at 30–31.

[39] SPITZER, *supra* note 13, at 17.

[40] *Terry*, 392 U.S. at 20, 23 ("It would have been poor police work indeed for an officer of 30 years' experience in the detection of thievery from stores in this same neighborhood to have failed to investigate this behavior further."); SPITZER, *supra* note 13 at 18; *see also* Akhil Reed Amar, Terry *and Fourth Amendment First Principles*, 72 ST. JOHN'S L. REV. 1097, 1098 (1998) ("Reasonableness—not the warrant, not probable cause—thus emerged as the central Fourth Amendment mandate and touchstone.").

[41] *Terry*, 392 U.S. at 20; SPITZER, *supra* note 13, at 18.

[42] United States v. Arvizu, 534 U.S. 266, 274 (2002); *see also* United States v. Sokolow, 490 U.S. 1, 7 (1989) ("We have held that probable cause means 'a fair probability that

The *Terry* Court celebrated Officer McFadden's experience in weighing his assessment of Mr. Terry's behavior. Subsequent courts have recognized the role experience and training play in officers' determinations of reasonable suspicion. "[The totality of the circumstances review] process allows officers to draw on their own experience and specialized training to make inferences from and deductions about the cumulative information available to them that 'might well elude an untrained person.'"[43] For instance, courts have deferred to an officer's 17 years of experience[44] and officers' past experience in similar situations[45] when crediting their findings of reasonable suspicion. Reasonable suspicion has been described as "commonsensical."[46] For the courts, whether a stop is constitutional is intertwined with an officer's experience and an underlying assumption that officers must be learning from their experiences.

B. STOP, QUESTION, AND FRISK IN NEW YORK CITY

1. Doctrinal Background

The constitutional bases for street stops by police in New York are codified by a state court case, *People v. De Bour*,[47] which elaborated on the reasonable suspicion standard for street stops created by the U.S. Supreme Court in *Terry v. Ohio*.[48] *De Bour* established a four-tiered framework to

contraband or evidence of a crime will be found,' . . . and the level of suspicion required for a *Terry* stop is obviously less demanding than that for probable cause . . . .") (citations omitted) (quoting Illinois v. Gates, 462 U.S. 213, 238 (1983)) (citing United States v. Montoya de Hernandez, 473 U.S. 531, 541 (1985)).

[43] *Arvizu*, 543 U.S. at 273 (quoting United States v. Cortez, 449 U.S. 411, 418 (1981)).

[44] United States v. Pack, 612 F.3d 341, 361 (5th Cir. 2010) ("[The officer's] suspicion is entitled to significant weight, because he had been a law enforcement officer for seventeen years."), *opinion modified on denial of reh'g*, 622 F.3d 383 (5th Cir. 2010).

[45] United States v. Timms, No. 17-CR-130 (KBF), 2017 WL 3503373, at *5 (S.D.N.Y. Aug. 16, 2017) (noting as evidence of reasonable suspicion that the officers "had both experienced situations involving multiple firearms in the past"); *see also* United States v. Sosunov, No. 17-CR-0350 (KBF), 2018 WL 2095176, at *1 n.1 (S.D.N.Y. May 7, 2018) (crediting the FBI agents' experience in "identifying, investigating, and dismantling criminal organizations").

[46] United States v. Lender, 985 F.2d 151, 154 (4th Cir. 1993) (explaining courts should "credit[] the practical experience of officers who observe on a daily basis what transpires on the street").

[47] De Bour v. People, 352 N.E.2d 562, 571–72 (N.Y. 1976).

[48] Terry v. Ohio, 392 U.S 1, 37 (1968). The term "reasonable suspicion" was introduced in dissent by Justice Douglas to contrast the majority's holding that accorded deference to the police officer's standardless judgment on what constitutes suspicious behavior. *Id.* at 20. ("The term 'probable cause' rings a bell of certainty that is not sounded by phrases such as 'reasonable suspicion.'").

govern police-citizen encounters in the state.[49] The *De Bour* case dealt with "whether or not a police officer, in the absence of any concrete indication of criminality, may approach a private citizen on the street for the purpose of requesting information."[50] This was distinct from and narrower than *Terry*'s deferential doctrine, which encoded broadly how officers applied their experience and judgment to decide that "criminal activity may be afoot."[51] Like the *Terry* Court, however, the New York Court of Appeals, the state's highest court, recognized the diversity of police work, noting that "[t]o consider the actions of the police solely in terms of arrest and criminal process is an unnecessary distortion. We must take cognizance of the fact that well over 50% of police work is spent in pursuits unrelated to crime."[52] That being so, police officers should be afforded "wide latitude to approach individuals and request information."[53]

De Bour's four tiers rise from least to most intrusive.[54] The third tier is a *Terry* stop.[55] Such a stop and detention is permitted when "a police officer entertains a reasonable suspicion that a particular person has committed, is committing or is about to commit a felony or misdemeanor," and officers have a corollary right "to frisk if the officer reasonably suspects that he is in danger of physical injury by virtue of the detainee being armed."[56] During the study period we analyze, when a NYPD officer conducted a stop, the officer filled out a UF-250 form to record information about the stop.[57] These

[49] *See* Spitzer, *supra* note 13, at 23–29; NYPD Patrol Guide, Procedure 212-11, Investigative Encounters: Right for Information, Common Law Right of Inquiry and Level 3 Stops 1–3 (2016), https://www.nyc.gov/html/nypd/downloads/pdf/analysis_and_planning/212-11.pdf [https://perma.cc/73D4-XAW4]; NYPD, Investigative Encounters Reference Guide 5 (2015), https://web.archive.org/web/20220126004339/http://nypdmonitor.org/wp-content/uploads/2016/02/InvestigativeEncountersRefGuideSept162015Approved.pdf [https://perma.cc/3HEY-DPDY].

[50] *De Bour*, 352 N.E.2d at 565.

[51] *Terry*, 392 U.S. at 30; *see also* Fagan, *supra* note 25, at 85.

[52] *De Bour*, 352 N.E.2d at 568 (citations omitted).

[53] *Id.* at 568.

[54] The first tier occurs when there is "minimal intrusion of approaching to request information [and] is permissible when there is some objective credible reason for that interference not necessarily indicative of criminality." *Id.* at 571–72. The second tier is the "common-law right to inquire." *Id.* at 572. Such a right "is activated by a founded suspicion that criminal activity is afoot and permits a somewhat greater intrusion in that a policeman is entitled to interfere with a citizen to the extent necessary to gain explanatory information, but short of a forcible seizure." *Id.* Finally, the fourth tier is an arrest, which may occur when an officer "has probable cause." *Id.*

[55] The resulting policy has been commonly termed stop and frisk, or stop, question and frisk (SQF).

[56] *De Bour*, 352 N.E.2d at 572.

[57] *See* SPITZER, *supra* note 13, at 89.

UF-250 forms provide a significant amount of data to researchers and litigators about NYPD's stop and frisk practices.[58]

2. *Floyd and the Stop and Frisk Litigation*

NYPD officers used their stop, question, and frisk authority widely,[59] provoking a lengthy and intensive litigation history.[60] Following the 1999 killing of an unarmed man by NYPD officers who claimed he resembled a suspected rapist, the New York Attorney General began an investigation into the NYPD's stop and frisk practices.[61] Also in 1999, class action plaintiffs filed suit in the Southern District of New York challenging the NYPD's stop and frisk practices, alleging the city "implement[ed] and enforc[ed], encourag[ed], and sanction[ed] a policy, practice and custom of unconstitutional stops and frisks."[62] In 2001, the court granted class certification, and in 2003, the City entered into a stipulated settlement with a class of plaintiffs that required the NYPD to establish a written policy on racial profiling and training programs, compile stop data using the UF-250 forms, and provide the data to class counsel.[63]

In 2008, plaintiffs again filed suit in federal court alleging the NYPD's stop and frisk practices were unconstitutional, in *Floyd v. City of New York*. In May 2012, the court granted class certification,[64] and following a bench trial, the court found the City "liable for violating plaintiffs' Fourth and Fourteenth Amendment rights."[65] At the same time, the judge ordered the City and NYPD to take certain remedial measures and appointed a monitor to oversee those measures and analyze the data produced.[66] NYPD stops and

[58] For instance, note the extensive use of the UF-250 data in the *Floyd* litigation. Floyd v. City of New York, 959 F. Supp. 2d 540, 559 (S.D.N.Y. 2013).

[59] *See* Tracey L. Meares, *The Law and Social Science of Stop and Frisk*, 10 ANN. REV. L. & SOC. SCI. 335, 339 (2014).

[60] Michael D. White & Henry F. Fradella, STOP AND FRISK: THE USE AND ABUSE OF A CONTROVERSIAL POLICING TACTIC 2 (2016); Meares, *supra* note 26, at 164–65.

[61] SPITZER, *supra* note 13, at 5, 9.

[62] Stipulation of Settlement at 1–2, Daniels v. City of New York, No. 99 Civ. 1695 (SAS) (S.D.N.Y. Sept. 24, 2003).

[63] *Id.* at 2–3, 5–11.

[64] Floyd v. City of New York, 283 F.R.D. 153, 160 (S.D.N.Y. 2012).

[65] Floyd v. City of New York, 959 F. Supp. 2d 540, 562 (S.D.N.Y. 2013).

[66] *Id.* at 563. The monitor is responsible for:

> [D]evelop[ing], in consultation with the NYPD and counsel for plaintiffs, a set of reforms of the NYPD's policies, training, supervision, auditing, and handling of complaints and discipline regarding stops and frisks and trespass enforcement. The monitor must also assess progress on the

frisks have fallen precipitously since the *Floyd* order was issued in 2013, but the racial distribution has remained relatively stable.[67]

3. Stop and Frisk in Practice

The court mandates for reporting on stop and frisk practices in New York City provided a rich dataset on police-citizen encounters and a window into contemporary urban policing.[68] Much of the research on stop and frisk in New York has focused on the racial disparities both in stops and their outcomes, on the use of force and other police-civilian interactions, and the crime control effects of those practices.[69] Across sampling, measurement, and analytic conditions, studies are more likely than not to show that minorities are disproportionately stopped compared to whites, and the efficiency, or hit rates, from these stops are lower than the hit rates for stops of whites.[70] The *Floyd* plaintiffs' expert reported that controlling for factors

> NYPD's implementation of these reforms and report to the court twice a year on the City's compliance with the court orders.

MARY JO WHITE, ROBERT L. CAPERS & BARBARA S. JONES, THE REPORT OF THE INDEPENDENT PANEL ON THE DISCIPLINARY SYSTEM OF THE NEW YORK CITY POLICE DEPARTMENT 15 (2019), https://www.independentpanelreportnypd.net/assets/report.pdf [https://perma.cc/56V6-YSCU]. To date, the monitor has produced eighteen reports. *See Resources & Reports*, NYPD Monitor, http://nypdmonitor.org/resource-reports [https://perma.cc/84BX-Z88X]. These reports include an analysis of NYPD stops. *See generally* PETER L. ZIMROTH, FIFTH REPORT OF THE INDEPENDENT MONITOR (2017) [hereinafter ZIMROTH, FIFTH REPORT], https://www.nyc.gov/assets/nypd/downloads/pdf/monitor-reports/2017-05-30-MonitorsFifthReport-AnalysisofNYPDStopsReported2013-2015-Asfiled.pdf [https://perma.cc/N8UP-ADB3]. Two other, related cases were also filed and settled along the same timeline. Davis v. City of New York, No. 10 Civ. 0699 (SAS) (S.D.N.Y. Jan. 28, 2010). *Davis* challenged stop and frisk policies stemming from vertical patrols and other "less formal sweeps" in New York City Housing Authority (NYCHA) buildings. *Id.* at 2. *Ligon v. City of New York*, No. 12 Civ. 2274 (SAS) (S.D.N.Y. Mar. 28, 2012), challenged stop and frisk policies "implemented pursuant to 'Operation Clean Halls,' a program that allows police officers to patrol inside and around thousands of private residential apartment building across [New York City]." *Id.* at 2.

[67] ZIMROTH, FIFTH REPORT, *supra* note 66, at 7–8.

[68] See Aziz Z. Huq, *The Consequences of Disparate Policing: Evaluating Stop and Frisk as a Modality of Urban Policing*, 101 MINN. L. REV. 2397, 2399 (2017).

[69] *See, e.g.*, Ben Grunwald & Jeffrey Fagan, *The End of Intuition-Based High-Crime Areas*, 107 CALIF. L. REV. 345, 347–54 (2019); John MacDonald, Jeffrey Fagan & Amanda Geller, *The Effects of Local Police Surges on Crime and Arrests in New York City*, 11 PLOS ONE 1, 1–3 (2016).

[70] *See* Jeffrey Fagan, *Recent Evidence and Controversies in "The New Policing"*, 36 J. POL'Y ANALYSIS & MGMT. 960, 964; GREG RIDGEWAY, ANALYSIS OF RACIAL DISPARITIES IN THE NEW YORK POLICE DEPARTMENT'S STOP, QUESTION, AND FRISK PRACTICES 40–42, (RAND Corp. 2007). *But see* Ridgeway, *supra*, at 13–19 (finding that Black pedestrians were not disproportionately stopped).

such as local racial composition, crime rate, and demographics, Blacks and Latinxs were more likely to be stopped than whites.[71] Separately, Fagan and colleagues found that "stops of whites are more 'efficient' and are more likely to lead to arrests, whereas those for [B]lacks and Hispanics are more indiscriminate," in that a greater percentage of stopped whites converted into arrests than stopped Blacks and Latinx persons.[72] Taking into account the heterogeneity of stops with respect to their bases of suspicion, two studies showed that stops which hewed closer to a probable cause justification had higher hit rates and were more likely to contribute to the security in local areas.[73]

One goal of stop and frisk is to reduce crime by disrupting planned or active crimes, and by deterring others through heightened risks of police encounters.[74] Whether stop and frisk actually leads to a decrease in crime is a contested claim.[75] Beyond any deterrent effects, studies have shown that very few stops result in any type of sanction, whether arrest or summons, or the recovery of weapons or contraband.[76] Moreover, once arrested, few stop and frisk arrests resulted in a conviction. "Close to half of all SQF arrests between 2009 and 2012 did not result in any conviction. Almost one in six arrests (15.7%) were never prosecuted."[77] Some studies have attributed stop

[71] *Floyd*, 959 F. Supp. 2d at 589.

[72] Andrew Gelman, Jeffrey Fagan & Alex Kiss, *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. Am. Stat. Ass'n 813, 820 (2007).

[73] Fagan, *supra* note 25, at 26, 84; MacDonald, Fagan & Geller, *supra* note 69, at 9–11.

[74] *See* Fagan, *supra* note 25, at 64 n.137 (noting the centrality of *Terry*'s crime control agenda); Meares, *supra* note 26,[60] at 165–69.

[75] *See generally* Nat'l Acads. of Sciences, Eng'g, Med., Proactive Policing: Effects on Crime and Communities (David Weisburd & Malay Majmundar eds., 2018), https://nap.nationalacademies.org/read/24928/chapter/1#ii [https://perma.cc/ZQQ2-VMNY]; Fagan, *supra* note 25, at 45–46; Meares, *supra* note 59, at 342–45. *Compare* MacDonald, Fagan & Geller, *supra* note, 69 at 10–11, *with* David Weisburd, Alese Wooditch, Sarit Weisburd & Sue-Ming Yang, *Do Stop, Question, and Frisk Practices Deter Crime? Evidence at Microunits of Space and Time*, 15 Criminology & Pub. Pol'y 31, 50 (2016).

[76] Sharad Goel, Justin M. Rao & Ravi Shroff, *Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy*, 10 Annals Applied Stat. 365, 375 (2016); Joseph Ferrandino, *The Efficiency of Frisks in the NYPD*, 2004–2010, 28 Crim. Just. Rev. 149, 162 (2012).

[77] Eric T. Schneiderman, N.Y. State Off. of the Att'y Gen., A Report on the Arrests Arising from the New York City Police Department's Stop-and-Frisk Practices 3 (2013). Additionally, 10.6% resulted in dismissal or acquittal, and another 21.3% resulted in an ACD (adjournment in contemplation of dismissal), a functional equivalent to dismissal. *Id.*

and frisk to the drop in New York City's crime rate, but the results tend to be ambiguous and contested.[78]

Constitutional compliance is a social good that reinforces the underlying moral norms of the law and its agents. Beyond the Fourth Amendment liability found in *Floyd*, empirical research has shown that police often disregard their constitutional mandates when stopping individuals. For example, an observational study conducted in a mid-sized city found substantial noncompliance with the Constitution.[79] Analyses of the 115 observed searches found that 30% were unconstitutional. Of the stop and frisk searches observed, 46% were unconstitutional. Monitoring of a police consent decree in Philadelphia shows that during the first five years of a consent decree,[80] the rate of Fourth Amendment non-compliance declined only slightly, from over 50% in the initial monitoring period in 2012 to 30% in 2018, six years after the initial report.[81]

C. IS THERE ROOM IN *TERRY* STOP PRACTICE FOR UPDATING?

We expect that police officers and agencies, on average, are rational actors that make decisions to maximize their benefits and minimize their costs or losses. In policing regimes, such as the NYPD's during the study period, where patrols are designed to proactively identify and interdict persons where "crime is afoot,"[82] we assume officers will be incentivized by supervisors and police executives to learn from their successes and failures and maximize the returns to crime control from everyday contacts.[83] These

[78] *See* Meares, *supra* note 5959, at 343–44; Christopher M. Sullivan & Zachary P. O'Keeffe, *Evidence that Curtailing Proactive Policing Can Reduce Major Crime*, 1 NATURE HUM. BEHAV. 730, 733 (2017).

[79] Jon B. Gould & Stephen D. Mastrofski, *Suspect Searches: Assessing Police Behavior Under the U.S. Constitution*, 3 CRIMINOLOGY & PUB. POL'Y. 315, 331–34 (2004); *see also* Grunwald & Fagan, *supra* note 69, at 350–52.

[80] The Consent Decree emerged from a lawsuit brought against Philadelphia on behalf of African American and Latino men who were stopped by police officers on the basis of their race or ethnicity. *See* Complaint ¶¶ 1–3, Bailey v. City of Philadelphia, No. 10-cv-5952-SD (E.D. Pa. Nov. 4, 2010); Settlement Agreement, Class Certification, and Consent Decree at 1, *Bailey*, No. 10-cv-5952-SD.

[81] Plaintiffs' Ninth Report to Court and Monitor on Stop and Frisk Practices: Fourth Amendment Issues at 1–4, *Bailey*, No. 10-cv-5952-SD.

[82] Terry v. Ohio, 392 U.S. 1, 30–31 (1968).

[83] *See* Charles F. Manski and Daniel S. Nagin, *Assessing Benefits, Costs, and Disparate Racial Impacts of Confrontational Proactive Policing*, 114 PROCEEDINGS NAT'L ACAD. SCIS. 9308–09 (2017) (developing a model of optimal policing based on tradeoffs between the social benefits and costs of proactive police tactics); Rachel A. Harmon & Andrew Manns, *Proactive Policing and the Legacy of* Terry, 15 OHIO ST. J. CRIM. L. 49, 56–57 (2017) (describing

positive returns include apprehending offenders, sanctioning violations, seizing weapons and contraband, and apprehending persons with outstanding warrants, as well as stops that produce good will or cooperation.

On patrol, whether in vehicles or on foot, officers are trained to proactively use their experience and judgment to identify possible crime suspects for attention and investigation.[84] Officers have limited time and resources to deter or prevent crime. Police stops that yield positive returns should signal to the officers the features of those stops that increase their reward or benefits. Having formed these beliefs, or priors, rational officers should observe the outcomes of their subsequent stops and form updated estimates of the probability of "good" stops. Officers should also learn the features of unproductive stops and adjust accordingly. A police officer acting rationally would pursue "good" stops and avoid unproductive ones by updating their beliefs through experience.

The reality of how stops are conducted provides a window into the updating process. Several studies have shown how officers form suspicion and decide to stop and possibly frisk an individual.[85] For example, "in three of four street stops in New York City, police observe a suspect for less than two minutes before proceeding to . . . an 'intrusion,'"[86] meaning officers make rapid decisions about their conduct. Applying psychological theories, researchers have investigated how police officers rely on "mental models" to inform future conduct.[87] Alpert and colleagues explain, "police officers learn to respond to people, places, and situations based on their experiences, including how they were trained and taught in the police academy, by field training officers, supervisors, and others."[88] These experiences can be racially biased or result in incorporating racially biased information into the officer's mental model.[89] Based on a study of police officers in Savannah, Georgia, Alpert and colleagues found that prior to making a stop, "[o]fficers were significantly more likely to form a non-behavioral suspicion when the

"[c]ontemporary proactive law enforcement" as intending to deter and prevent crime, rather than to uncover or directly stop it).

[84] *See* Harmon & Manns, *supra* note 83, at 55–58.

[85] For a review of the literature, see Geoffrey P. Alpert, John M. MacDonald & Roger G. Dunham, *Police Suspicion and Discretionary Decision Making During Citizen Stops*, 43 CRIMINOLOGY 407, 409–11 (2005); Jeffrey Fagan & Amanda Geller, *Following the Script: Narratives of Suspicion in* Terry *Stops in Street Policing*, 82 U. CHI. L. REV. 51, 56–61 (2015).

[86] Fagan & Geller, *supra* note 85, at 63.

[87] *Id.* at 65.

[88] Alpert, MacDonald & Dunham, *supra* note 85, at 413–14.

[89] *Id.*

suspect is [B]lack,"[90] and "the longer the officers had been on the police force, the more likely they were to form non-behavioral suspicions."[91] Perhaps as officers gain more experience, they may form certain mental models or schemas of suspicion.[92] Similarly, Geller and Fagan found that "officers defaulted to convenient and stylized narratives to justify stops," and sidestepped the individualized suspicion requirement of the Fourth Amendment.[93] In terms of other related, officer-based research, there have been studies into "working memory capacity,"[94] emotion and anger,[95] and the sequential decisions during a specific incident.[96] Other research has investigated organizational culture and learning[97] and modes of learning from others.[98] Finally, experimental research has shown that police officers tend to be over-confident in their ability to detect lies or suspicious activity.[99]

Other studies illustrate the potential for carefully articulated and prudently applied bases of suspicion to improve the search for weapons and

[90] *Id.* at 422. "Nonbehavioral criteria included officer concern about an individual's appearance, the time and place, and descriptive information provided to an officer. Suspicions based on nonbehavioral criteria do not necessarily provide a clear justification for a stop." *Id.* at 419.

[91] *Id.* at 422.

[92] *Id.* at 422–33.

[93] Fagan & Geller, *supra* note ^{85}85, at 86.

[94] Defined as the capacity for executive control (the extent to which individuals can exert control over their decision-making processes). Heather M. Kleider, Dominic J. Parrott & Tricia Z. King, *Shooting Behaviour: How Working Memory and Negative Emotionality Influence Police Officer Shoot Decisions*, 24 APPLIED COGNITIVE PSYCH. 707, 715 (2010) (showing that among police officers, lower working memory capacity was associated with a greater likelihood of shooting unarmed targets and a failure to shoot armed targets).

[95] Shanique G. Brown & Catherine S. Daus, *The Influence of Police Officers' Decision-Making Style and Anger Control on Responses to Work Scenarios*, 4 J. APPLIED RSCH. MEMORY & COGNITION 294, 294 (2015).

[96] Lorie A. Fridell & Arnold Binder, *Police Officer Decisionmaking in Potentially Violent Confrontations*, 20 J. CRIM. JUST. 385, 385–86 (1992).

[97] *See, e.g.*, Barry Sugarman, *Organizational Learning and Reform at the New York City Police Department*, 46 J. APPLIED BEHAV. SCI. 157, 157 (2010).

[98] *See, e.g.*, Allison T. Chappell & Alex R. Piquero, *Applying Social Learning Theory to Police Misconduct*, 25 DEVIANT BEHAV. 89, 89 (2004); Anja J. Doornbos, Robert-Jan Simons & Eddie Denessen, *Relations Between Characteristics of Workplace Practices and Types of Informal Work-Related Learning: A Survey Study Among Dutch Police*, 19 HUM. RES. DEV. Q. 129, 129 (2008); Johan Lundin & Urban Nuldén, *Talking About Tools – Investigating Learning at Work in Police Practice*, 19 J. WORKPLACE LEARNING 222, 222 (2007).

[99] Michael G. Aamodt & Heather Custer, *Who Can Best Catch a Liar? A Meta-Analysis of Individual Differences in Detecting Deception*, 15 FORENSIC EXAM'R 6, 10 (2006); Eugenio Garrido, Jaume Masip & Carmen Herrero, *Police Officers' Credibility Judgments: Accuracy and Estimated Ability*, 39 INT'L J. PSYCH. 254, 256 (2004).

other contraband.[100] By using an ex-ante probability that a stop will successfully conclude with recovery of contraband, officers' hit rates might improve if the known factors that contributed to success could be signaled or somehow communicated to officers conducting stops.[101] Goel and colleagues developed such a model using the *Floyd* data to identify characteristics of successful stops, which included a full range of information about suspects, locations and times, and the same bases of suspicion that we use in this project.[102] Their model was based on a test data set of stops from 2008–2010, using random pairs of stops that produced alternate outcomes. The model's accuracy rate was 83%, a far cry from the 0.1% hit rate of NYPD officers overall in recovering weapons from 2004–2009.[103] They, in effect, designed an algorithm to predict hit rate success for weapons using a rich set of information based on officers' prior stop outcomes. This method is very much in line with the Bayesian approach that we use in the second analysis in this Article. Where we depart is the use of officer-level data to identify whether officers use a similar algorithm to learn the criteria of successful hits through a learning-updating exercise that the Goel and colleagues' SHR (stop-level hit rate) model produced.

Few studies have considered on-the-ground stop and frisk rates of individual officers.[104] Of those that have, they have found problematic behavior—both inaccuracy and constitutional errors—tends to be "highly concentrated in a few officers."[105] Also, few empirical analyses have considered whether officers learn or improve over time. An analysis of searches by Florida highway patrol officers from 2000 to 2001 indicated that "[t]he same officers who succeeded [by recovering evidence in probable cause searches] in 2000 also succeeded at high rates in 2001. Similarly, officers who were less successful in 2000 tended to be relatively less successful in 2001."[106] At the precinct level and in terms of stops and frisks yielding weapons, research has shown mixed results: "Directly comparing the first year (2004) [efficiency] score with the last year (2010), 35 precincts

[100] Goel et al., *supra* note 11, at 181.

[101] *Id.* at 211–12; *see also* Shroff, *Statistical Tests*, *supra* note 100, at 567.

[102] Goel et al., *supra* note 11, at 212–13.

[103] Floyd v. City of New York, 959 F. Supp. 2d 540, 559 (S.D.N.Y. 2013) (noting that the rate of gun seizures was 0.1%).

[104] For exceptions, see Antonovics & Knight, *supra* note 10, at 163; Billie R. Close & Patrick L. Mason, *Searching for Efficient Enforcement: Officer Characteristics and Racially Biased Policing*, 3 REV. L. & ECON. 263, 265 (2007); Jeffrey Fagan, Anthony A. Braga, Rod K. Brunson & April Pattavina, *Stops and Stares: Street Stops, Surveillance, and Race in the New Policing*, 43 FORDHAM URB. L.J. 539, 540 (2016); RIDGEWAY, *supra* note 70, at 1.

[105] Gould & Mastrofski, *supra* note 7379, at 344; RIDGEWAY, *supra* note 7070, at 28.

[106] Minzner, *supra* note 4, at 931.

had lower efficiency scores, . . . 10 showed no change, . . . and 31 showed an increase in efficiency score."[107] Both studies compare rates based on the aggregation of stop outcomes over a relatively long time horizon, often too long to fully assess learning and improvement from event to event over time. Officers may learn from successes in the past month but forget the conduct that led to success six months ago. In those cases, officers may be more likely to view a situation anew and decide unconditioned by prior decisions.

From the perspective of learning or updating processes, it is not entirely clear how officers learn from and how they evaluate their own success. First, informal on-the-job learning depends on how often officers find themselves in the same or similar situations. If officers routinely encounter similar situations and can deploy the same skills, past experience can be informative. On the other hand, if officers are rarely in similar situations in which they can exercise discretion, "the kinds of skills that experience teaches are less helpful."[108] However, past experience may assist in the ability to adapt to new experiences and make decisions.[109] Additionally, it is unclear how officers process their past experiences. For instance, "[p]olice may . . . suffer from inappropriately extrapolating from past results when they have insufficient information to identify a trend or an important factual distinction."[110] Officers, even those with significant experience, will benchmark and generalize from a small sample of data; hindsight and confirmation bias can cloud officers' recollections of what made a specific event suspicious. Additionally, officers rarely received feedback on past accuracy rates.[111]

Police officers may have a hard-to-quantify understanding of success. They may see success as "avoiding affronting the department or getting seriously hurt or sued," rather than "intelligently discriminating in their tactical choices so that they are raising the probability of achieving stated goals."[112] Moreover, if officers cannot distinguish "lucky" searches from those brought about by reasoned decision-making, they may be stymied in their efforts to identify and isolate patterns of suspicious behaviors. In fact,

[107] Ferrandino, *supra* note 76, at 154.

[108] David H. Bayley & Egon Bitter, *Learning the Skills of Policing*, 47 L. & CONTEMP. PROBS. 35, 38 (1984).

[109] Michael L. Birzer & Robert E. Nolan, *Learning Strategies of Selected Urban Police Related to Community Policing*, 25 POLICING 242, 249 (2002).

[110] Andrew E. Taslitz, *Police Are People Too: Cognitive Obstacles to, and Opportunities for, Police Getting the Individualized Suspicion Judgment Right*, 8 OHIO ST. J. CRIM. L. 7, 44 (2010).

[111] *See* Garrido, et al., *supra* note 99, at 267–68.

[112] Bayley & Bitter, *supra* note 108, at 47.

officers' perceptions of good search results vary widely.[113] These individualized learning models may well get in the way of an institutional response that would promote positive outcomes and improvements in officers' accuracy over time.

Recently, work by Tracey Meares and others has highlighted the extent to which stop and frisk is less of a tool utilized by officers investigating a single incident of ongoing misconduct and more of a program "carried out by a police force en masse."[114] Doctrinally, understanding stop and frisk as a widespread policy and program presents challenges. *Terry* and Fourth Amendment law's focus on the individual incident at hand has created a "mismatch between the level at which the Supreme Court articulated the relevant test . . . and the scale at which police today (and historically) engage in stop and frisk as a practice."[115] Officer conduct is reviewed at the incident level, but officers carry out stop and frisk at a programmatic level, and important conduct that informs programmatic level decisions may not appear relevant when narrowing the focus to a single incident. It also presents challenges to an officer's ability to learn. As stopping moves away from "a collection of individual investigations occurring between an officer and a person that the officer believes to be committing a crime" and toward directives to programmatically engage in preventative patrols, where the mere making of stops is seen as beneficial,[116] there is less and less incentive for officers to engage in critical thinking about which stops resulted in weapons recovery or arrest and to apply that knowledge going forward. If an officer's task is simply to maximize the number of stops made, officers have little reason to focus on indicators of present misconduct, but rather concentrate on "policing people they suspect *could* be suspects"[117]—which gets very close to the reliance on hunches the *Terry* Court sought to avoid. Increased time on the job may not parallel better results (in terms of weapons recovery and arrests) in these cases. If officers continue to make unproductive stops without consequence and fail to understand what constitutes valid reasonable suspicion, the risk of constitutional error grows.

[113] *See, e.g.*, PAUL QUINTON, NICK BLAND & JOEL MILLER, HOME OFF., POLICE STOPS, DECISION-MAKING AND PRACTICE 53–61 (Joel Miller ed., 2010); William F. Walsh, *Patrol Officer Arrest Rates: A Study of the Social Organization of Police Work*, 3 JUST. Q. 271, 286–88 (1986).

[114] Meares, *supra* note 26, at 162–63; *see also* Meares, *supra* note 59, at 340; Goel et al., *supra* note 11, at 189–90.

[115] Meares, *supra* note 26, at 162.

[116] *Id.* at 162, 171–72; *see also* Meares, *supra* note 59, at 340 (describing the NYPD's policies as "a planned and concerned effort to drive crime down rather than intervening in crimes in progress").

[117] *Id.* at 164.

D. FREQUENTIST AND BAYESIAN TESTS OF OFFICER LEARNING AND UPDATING

In this study, we track officers over time to understand whether and to what extent NYPD officers improve the accuracy of their decisions to stop individuals on the street. Whereas prior research has focused on the current decision to stop vis-à-vis factors such as race and neighborhood, we consider whether past stops and their outcomes inform the outcome of future stops. We extend this work by disaggregating stops according to the basis of suspicion, the approximate constitutional threshold of the stop (probable cause versus the lower standard of reasonable suspicion), and the suspected crime. Additionally, we look at a narrower time frame of one month for learning compared to the longer intervals considered in other studies.

Specifically, we use two different analytic and conceptual frameworks. We first apply a frequentist methodology—an approach based on a repeated sampling of an underlying population, which "centers on the question of how unlikely it would be to observe the actually-observed value in the counterfactual event that the variable of interest had a particular value." [118] A frequentist analysis permits us to engage in traditional null hypothesis testing—examining in this analysis whether there is a statistically significant relationship between the current month's hits and those of prior months. Considering the number of successful stops in a month ("hits"), we use random effects Poisson regressions to consider whether the relationship between officers' success in the current month and successful stops in each of the prior three months. We then pivot to a Bayesian approach—an approach concerned with combining prior knowledge about parameters with the observed data to form a posterior distribution, and with "how new information should cause a person to update her beliefs about the probability that a proposition is true."[119] This approach allows us to examine the extent to which accounting for prior probabilities influences the analysis of hits.

In this case, a Bayesian analysis allows us to view updating through a perspective that considers simultaneously the probabilities of "guilt" and the "innocence." Prior to making a stop, an officer making a stop may calculate the probability of the suspect being "guilty," but the probability remains that the suspicious behavior was entirely innocent and explainable by a host of other circumstances.[120] This prior probability is an outcome that we might

[118] Jonah B. Gelbach, *Estimation Evidence*, 168 U. PA. L. REV. 549, 561 (2020).

[119] *Id.* at 564.

[120] E. T. JAYNES, PROBABILITY THEORY: THE LOGIC OF SCIENCE 243 (G. Larry Bretthorst ed., 2003) (noting that some theories may not work "acceptably unless other circumstances are present").

expect based on our experiences and observations of the antecedents of criminal activity (the base rate), informed by what our prior experiences and background information tells us about similarly situated events.[121] We apply a similar Poisson regression structure to our Bayesian analysis as with our frequentist regressions, relying on uninformative default priors, to calculate the probability of successful stops in the current month based on prior months' stop outcomes and a host of other factors.

Some real-world examples provide context for these two learning models. Consider *Ornelas v. U.S.*, where a loose door panel and a rusty screw in a 10-year-old car were the bases for an inference by police that the car was being used to smuggle drugs.[122] Ornelas' name had appeared on a registry of known heroin dealers. Officers concluded that the car was carrying smuggled heroin rather than concluding that it simply was old and needed repair. The officers searched the car and found the contraband. This probabilistic assessment gave rise to the search and arrest. But had they been wrong, and the car was merely in need of repair, how would the same officers regard the next instance with similar parameters?

Similarly, why did Jimmy Warren run from the Boston police when the police attempted to stop and question him?[123] Police might have thought that Warren ran because he was evading arrest for criminal activity. Or Warren could have fled because he was carrying a gun. But, as the court concluded, he also might have fled out of fear of a violent confrontation with police.[124] The court found his fear of the police to be grounded in facts and reasonable probability, an implicit Bayesian view of Warren's behavior, where the court invoked background facts and contextual factors.[125] Compare this with

[121] W. David Ball, *The Plausible and the Possible: A Bayesian Approach to the Analysis of Reasonable Suspicion*, 55 AM. CRIM. L. REV. 511, 513–15 (2018) ("Rather than start with an explanation and evaluate the likelihood of observing behaviors consistent with that explanation, a Bayesian approach would start with the data and ask which explanation is more plausible.").

[122] Ornelas v. United States, 517 U.S. 690, 693–94 (1996).

[123] Commonwealth v. Warren, 58 N.E.3d 333, 337–38, 341–43 (Mass. 2016).

[124] The court stated:

> [I]n such circumstances, flight is not necessarily probative of a suspect's state of mind or consciousness of guilt. Rather, the finding that black males in Boston are disproportionately and repeatedly targeted for FIO encounters suggests a reason for flight totally unrelated to consciousness of guilt. Such an individual, when approached by the police, might just as easily be motivated by the desire to avoid the recurring indignity of being racially profiled as by the desire to hide criminal activity. Given this reality for black males in the city of Boston, a judge should, in appropriate cases, consider the report's findings in weighing flight as a factor in the reasonable suspicion calculus.

Id. at 540.

[125] *Id.*

Illinois v. Wardlaw, where, when faced with a defendant who fled, the Supreme Court noted that while "there are innocent reasons for flight from police," that "does not establish a violation of the Fourth Amendment" since lawful conduct can be indicative of criminal activity.[126] Like Warren, Mr. Wardlow might also have been speeding through a police-saturated neighborhood fearing a confrontation and an arrest. The problem in these cases is that reasonable suspicion, the legal standard and theory, requires that criminal activity be the most likely explanation. A Bayesian might weigh that explanation against competing explanations, particularly after aggregating knowledge from recent civilian interactions, and then decide whether the theory is plausible, including background factors in the analysis as well as how often the theory itself is wrong.

A narrow legal and frequentist framework would attach the same weights to the facts, given the totality of the circumstances giving rise to suspicion. But a Bayesian would reconsider the probability attached to the original set of circumstances, creating a space for the possibility of the alternate explanation, and estimate the probability based on prior estimates that informed their confidence in the theory. "Bayesians start with the data and then fit the theories; frequentists start with theories and then fit data."[127] The Bayesians would fit the theory to the facts, while the Frequentist would condition the results on a hypothesis (i.e., a theory) and estimate the probability that the hypothesis is correct. We adopt these distinctions in assessing whether police update their use of reasonable suspicion to affect an investigative stop.

II. DATA AND METHODS

A. DATA AND SAMPLE

We analyze data provided by the NYPD to address the question of whether officers' accuracy changes and improves over time. The dataset contains de-identified records for each stop and frisk event, or *Terry* stop, made by the NYPD during the study period, January 2004 to December 2016, as recorded on UF-250 forms.[128] In addition to the publicly available stop data, these records included an encrypted officer identifier for each stop. We created a novel dataset in which individual officers can be tracked over time

[126] Illinois v. Wardlaw, 528 U.S. 119, 125–26 (2000).

[127] Ball, *supra* note 121, at 518.

[128] THE STOP, QUESTION AND FRISK DATA, NYC OPENDATA, https://data.cityofnewyork.us/Public-Safety/The-Stop-Question-and-Frisk-Data/ftxv-d5ix [https://perma.cc/7KQZ-XBU6]. In addition, geocodable data were provided to the *Floyd* plaintiffs on stops, crime complaints, and arrests.

across stops. Unfortunately, the data do not contain information about officer demographics.[129]

The data include information that allowed us to measure the stop "career" of each officer, including the officer's first stop in any year of the study period. Officers only appear in the data if they made one or more stops, so an officer who is patrolling but records no stops is not included. In order to form the panel, each officer is arrayed by her own "career." Rather than associating each stop with the calendar year of the stop, the stop is associated with the year of the officer's career in which it is made, based on the year that each officer first appears in the data. For example, if an officer first made a stop in 2009, that would be listed as Year 0 for that officer; similarly, if another officer's first stop was in 2012, 2012 would be listed as Year 0 for that officer. The panel, therefore, aligns officers based on their experience and presence in the dataset, rather than the year of their stop. This permits us to create a panel of officers more directly tied to experience with or exposure to making stops. We do, however, also control for the calendar year of each stop, in order to control for the unique factors within the specific year, as well as the month tenure of each officer over her stop career. In order to ensure we are not considering officers who made a substantial number of stops prior to the start of the study period in 2004, we exclude from the panel officers whose first stop was in 2004 or 2005—only those officers who first appear in 2006 and onward are included. Additionally, we exclude from the panel those officers whose first stop is in 2016 in order to allow for at least one full year of tenure in the panel. The result is a database of 17,900 officers with a total of approximately 2.05 million stops from 2006–2016 in all five boroughs (counties) in the city.[130]

[129] Though the data do not contain officer demographics, several studies indicate that officer demographics play a role in the demographics of the individuals stopped. *See, e.g.*, Antonovics & Knight, *supra* note 10, at 163 (finding that Boston police are more likely to search a driver if the race of the officer differs from the driver); Close & Mason, *supra* note 104, at 301 (finding white Florida Highway Patrol officers "display bias in the propensity to search African American and Latino drivers"); Jeffrey Fagan, Anthony A. Braga, Rod K. Brunson & April Pattavina, *Stops and Stares: Street Stops, Surveillance, and Race in the New Policing*, 43 FORDHAM URB. L.J. 539, 540 (2016) (showing that relative to white suspects, Black suspects are more likely to be observed, interrogated, and frisked or searched controlling for gang membership and prior arrest history). These studies, however, were conducted in areas with substantially lower quota pressure than exerted by the NYPD during the study period. Accordingly, we suspect that these differentials are tempered by the strict enforcement of quotas. *C.f.* Nathaniel Bronstein, Note, *Police Management and Quotas: Governance in the CompStat Era*, 48 COLUM. J.L. & SOC. PROBS. 543, 550 (2014) ("Quotas . . . restrict discretion.").

[130] King County (Brooklyn), Queens County, Richmond County (Staten Island), Bronx County, New York County (Manhattan).

We also include a discontinuity marking the May 2012 date of the class certification in the *Floyd* litigation in New York City, a point when the NYPD began a precipitous drop in stops as the litigation advanced to a period of intensive pretrial activity followed by the trial a year later, leading to a court order and supervision at the end of 2013.[131] We estimate the effects of the *Floyd* litigation using this marker and present additional analyses based on that date.

The dataset includes a wide range of variables, with demographic information about the age, gender, and race or ethnicity of the individual stopped, the suspected crime, the bases of suspicion motivating the stop, whether a frisk or search ensued, whether force was used, the duration of the stop, and the stop outcome (arrest made or summons issued, seizures of weapons or contraband). The data also include information on the address where the stop was made and the work assignment (command) of the officer making the stop.

B. MEASURES

1. *"Hit Rates"*

We measure changes over time in two different configurations of officers' "hits": (1) those stops resulting in an arrest or summons and (2) those resulting in the recovery of weapons or contraband. We then break down the latter measure into separate outcomes: weapons seizures and contraband seizures. These rates were calculated on a monthly basis for each officer, and then lagged by one, two, and three months.[132] Outcomes were estimated controlling for several contextual stop-related covariates: suspect race, suspected crime,[133] stop actions (i.e., frisk or search), the basis of reasonable suspicion, and the borough where the stop was made.

[131] Floyd v. City of New York, 283 F.R.D. 153, 164 (S.D.N.Y. 2012); *see also* Mummolo, *supra* note 10, at 4 (identifying several shocks that began with the class certification order and continued through the conclusion of the *Floyd* litigation).

[132] For example, if an officer made arrests in June, July, August, and October 2006, lagged hit rates will be created for July (June's rate) and August (July's rate), but not for June or October, as there were no stops made in the previous month.

[133] Report of Jeffrey Fagan, Ph.D at 50–55, Floyd v. City of New York, 959 F. Supp. 2d 540 (S.D.N.Y. 2013) (No. 08 Civ. 1034) [hereinafter Fagan, Expert Report]. Stop data also included police reports of the crime suspected in each stop. These included over 150 specific codes that were reduced to seven categories that reflected the categories of interest in the policy debate on crime in New York. *Id.* at 9.

The validity of hit rates as a measure of police accuracy in street and highway stops is contested, to say the least.[134] In both highway and street stops, early hit rate analyses assumed that drivers of different races would adjust their propensity to carry drugs or other contraband based on their probability of being detected in a stop, eventually reaching an equilibrium where hit rates are equalized even if the stop and search rates are skewed.[135] This construction raises several questions, both empirically and constitutionally. First, an equilibrium model on hit rates assumes that a propensity for individuals to carry weapons or contraband is knowable, and without obvious markers of suspicion, can only be inferred actuarially or historically from group properties and ecological variables. This is tantamount to actuarial or group suspicion, or more commonly, profiling. That assumption of collective suspicion is a frequentist inference that is indexed to other crime measures associated with a place or a group: reported crimes, arrest patterns, neighborhood composition, or, in the case of firearms, victim reports of suspect characteristics in robberies or shootings. If the hit rate for minorities who are stopped at a higher rate is lower than for whites, then police may be applying a lower standard of suspicion—rooted in these group markers—for non-white suspects in deciding whom to stop or search.[136] In that case, learning or updating is constrained when a narrow set of biased—or at the least weakly predictive—markers inform the officer's priors.

This turned out to be the case in the *Floyd* litigation, leading the trial judge to charge the police with a policy and practice of "indirect profiling."[137] Others have criticized hit rate tests as burdened by both omitted and included variable biases,[138] where designs either omit non-racial explanations of discrimination or overload models with race-correlated variables that do not offer legitimate reasons for unjustified actions. Inframarginality in hit rate

[134] *See* Ayres, *supra* note 9, at 135, 138–140; Knowles, Perisco & Todd, *supra* note 9, at 214, 223; *see also* Simoiu et al., *supra* note 9, at 1193–94 (claiming the inaccuracy of between group comparisons when the groups have different risk distributions).

[135] Knowles et al., *supra* note 9, at 205–06.

[136] Ridgeway & MacDonald, *supra* note 11, at 9 (hypothesizing that differential hit rates by race suggest that officers apply a different standard for the population group with higher search rates but lower hit rates).

[137] Floyd v. City of New York, 959 F. Supp. 2d 540, 562 (S.D.N.Y. 2013).

[138] Ian Ayres, *Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of "Included Variable" Bias*, 48 PERSP. BIOLOGY & MED. 568 (2005); *see also* Clifford C. Clogg & Adamantios Haritou, *The Regression Method of Causal Inference and the Dilemma Confronting This Method*, *in* CAUSALITY IN CRISIS? STATISTICAL METHODS AND THE SEARCH FOR CAUSAL KNOWLEDGE IN THE SOCIAL SCIENCES 83–112 (Vaughn R. McKim & Stephen P. Turner eds., 1997).

tests also distorts conclusions about group effects. While the concerns over these biases have focused on disparate impact by race, they generalize to the use of outcomes as a measure of—in this case—police decision making in the conduct of investigative stops.[139]

2. *Reasonable Suspicion and Probable Cause*

The basis of suspicion for each stop was categorized as approximating probable cause or reasonable suspicion, two different thresholds defined in constitutional case law to justify police intrusions; stops were classified accordingly.[140] To indicate the basis of suspicion, police were provided a set of checkboxes to record their legal justification for each stop.[141] The boxes included affirmative stop rationales plus an option to check "other" and record the specifics by hand.[142]

[139] Despite the limitations of stop outcomes as metrics to evaluate the constitutionality and efficacy of investigative stops, we use outcomes, or hit rates, to assess updating among police officers. We do so for both empirical and practical reasons. First, these are the most relevant features of policing to gauge officer choices and actions with respect to involuntary encounters with civilians. Unreported or poorly documented stops pose the risk of measurement error, but we also note the institutional incentives to record stops to fulfill mandates to demonstrate police activity. Second, hit rates provide a measure of the normative component of policing: the burden on the innocent who are stopped with no evidence of wrongdoing, and the ensuing dignitary harms that ensue from the deprivations of liberty and the exposure of those stopped to police violence. The risk of "petty indignity" was a concern of the *Terry* court, but not enough of a concern to refrain from delegating discretion to police as to which behaviors were indicia of suspicion and what to do once encountered. *See* Terry v. Ohio, 392 U.S. 1, 16–18 (1968) ("Moreover, it is simply fantastic to urge that such a procedure performed in public by a policeman while the citizen stands helpless, perhaps facing a wall with his hands raised, is a 'petty indignity.'"); *see, e.g.*, Josh Bowers, *Probable Cause, Constitutional Reasonableness, and the Unrecognized Point of a Pointless Indignit*y, 66 STAN. L. REV. 987, 991 (2014); Devon W. Carbado & Jonathan Feingold, *Rewriting* Whren v. United States, 68 UCLA L. REV. 1678, 1686 (2021) (characterizing the burden on the innocent as sacrificing privacy, dignity, and security for the "greater good"—a sacrifice that others are never asked, nor expected, to bear. That sacrifice can only be considered the "greater" good if you do not account for those experiencing the harm"). *See generally* William J. Stuntz, Terry*'s Impossibility*, 72 ST. JOHN'S L. REV. 1213 (1998) (citing a range of potential harms to the suspect who is stopped but who has broken no law).

[140] *See Terry*, 392 U.S. at 28; Fagan, *supra* note 25, at 48 (discussing the Supreme Court's departure from the certainty of probable cause toward a more capacious reasonable suspicion standard that could justify both street stops and protective frisks)51.

[141] The checkboxes were incorporated into the standard reporting form for stops, the UF-250. They were a set of indicia of suspicion derived from the aggregate experiences of officers who had been conducting stops over many years. *See* Fagan, Expert Report, *supra* note 133, at 48–49; Fagan, *supra* note 25, at 68; Fagan & Geller, *supra* note 85, at 68–69.

[142] *See* Fagan, *supra* note 25, at 68. In about 95 percent of the stops from 2004 – 2016, officers checked from one to six factors, creating 60,459 possible combinations that express the bases of suspicion for this subset. *Id.*

Following MacDonald et al.[143] and Fagan,[144] we organized these stops into nine categories of suspicion that incorporated a set of behavioral categories based on both state and federal case law that would survive a Fourth Amendment test for the individualized stop rationales.[145] Three of the nine factors describe observable suspect behaviors that approximate criminal activity: (1) actions indicative of engaging in drug transactions; (2) actions indicative of violent crimes; or (3) "casing" victim or location.[146] Each factor is narrow and behaviorally specific, avoiding the vagueness and subjectivity that worried the *Terry* court.[147] The behavioral grounding of these three categories provides little room for cognitive error or perceptual distortion, and are consistent with state and federal case law on probable cause.[148] In addition, courts have said that observed criminal behaviors are sufficient on their own to justify a police stop.[149]

We used these three categories of stop rationales to sort stops into *probable cause* versus *reasonable suspicion* stops. By separating out three categories of suspicion that are closer in meaning to a probable cause standard, the empirical strategy assessed whether stop outcomes, or hit rates, varied according to the restrictiveness of the suspicion threshold for the stop. We estimate the number of probable and non-probable cause stops in each month to assess their separate and combined effects on subsequent stop outcomes. We would assume that the more precise information content in probable cause stops would inform and improve the learning and updating byproducts of these interactions.

In contrast, the other six bases or categories of suspicion require subjective judgments and attributions of intent: (1) furtive movements, (2) fits descriptions, (3) carrying objects in plain view, (4) suspicious bulge, (5) evasive actions, or (6) "other."[150] In contrast to observations of specific

[143] MacDonald et al., *supra* note 69.

[144] Fagan, *supra* note 25.

[145] Fagan, Expert Report, *supra* note 133, at app. D (listing suspicion categories); *see also* Adams v. Williams, 407 U.S. 143, 146 (1972) ("A brief stop of a suspicious individual, in order to determine his identity or to maintain the status quo momentarily while obtaining more information, may be most reasonable in light of the facts known to the officer at the time.").

[146] *See* Terry v. Ohio, 392 U.S. 1, 28 (1968); United States v. Padilla, 548 F.3d 179, 187–88 (2d Cir. 2008); People v. Richard, 668 N.Y.S.2d 386, 387 (N.Y. App. Div. 1998).

[147] *See* William J. Stuntz, Terry*'s Impossibility*, 72 ST. JOHN'S L. REV. 1213, 1215–16 (1998); Fagan & Geller, *supra* note 85, at 52–54.

[148] Fagan, Expert Report, *supra* note 133, at app. D.

[149] Floyd v. City of New York, 959 F. Supp. 2d 540, 566–67 (S.D.N.Y. 2013).

[150] Fagan & Geller, *supra* note 85, at 71. "Other" stop factors were checked off at frequencies that varied by type of suspected crime. The text strings for the "other" factor were

criminal activity in probable cause stops, these subjective factors are vulnerable to cognitive bias and error, as well as racialized attributions of suspicion or criminality.[151] They would provide little choate or substantive information for internalization and learning about success and failure. One simple reason is that if these bases of suspicion were subjective and essentially "hunches," the learning residual from these stops would be both minimal and well below the learning threshold of the more cognitively precise probable cause stops.

C. EMPIRICAL STRATEGY

We address two related "critical question[s]": "[D]o success rates of officers vary and are they consistent over time? Is an officer's history of success or failure in the past predictive of her probability of success in the future?"[152] We begin with descriptive statistics on stops and outcomes. We then use regression analyses to examine officer behavior. The dataset is collapsed on a unique officer ID, so that each officer's monthly performance can be compared over time, forming a panel.

The first set of analyses are random effects Poisson regressions using panel models that simply lag the prior month's hits for each officer on the current month. We use the number of stops of each officer in the current period as the exposure for hits, effectively converting the count of hits into a rate. We use random effects owing to the uncertain sampling distributions of officers, and some uncertainty in officer compliance with reporting requirements. We model hits resulting in arrests or summonses; weapons or contraband; weapons only; and contraband only. Standard errors are clustered at the officer level. The model takes the general form:

$$\delta_{it} = \beta 0_{it} + \beta_1 HitRate_{it-1} + \beta_2 ReasonableSusp_{it-1} + \beta_3 StopRationale_{it-1} + \beta_3 TotStops_{it-1} + \beta_4 Race_{t-1} + \alpha_i + u_{it},$$

a diverse set of observations that were at times specific (e.g., smell of marijuana smoke) and at times bizarrely vague (e.g., looks like a perp). Second Supplemental Report of Jeffrey Fagan, Ph.D. at 27–28, app. C, Floyd v. City of New York, 959 F. Supp. 2d 540 (S.D.N.Y. 2013) (No. 08 Civ. 1034); *Floyd*, 959 F. Supp. 2d at 559.

[151] *See, e.g.*, Alpert, MacDonald & Dunham, *supra* note 85, at 422–23; Adam M. Samaha, *Regulation for the Sake of Appearance*, 125 HARV. L. REV. 1563, 1620–34 (2012); Robert J. Sampson, *When Things Aren't What They Seem: Context and Cognition in Appearance-Based Regulation*, 125 HARV. L. REV. F. 97, 99–102 (2012) [hereinafter Sampson, *When Things Aren't*]; Robert J. Sampson & Stephen W. Raudenbush, *Seeing Disorder: Neighborhood Stigma and the Social Construction of "Broken Windows"*, 67 SOC. PSYCH. Q. 319, 330–34 (2004).

[152] *See* Minzner, *supra* note 4, at 930.

where δ_{it} is the count of the current month's hits over the study period and $\beta_1$is the past months' where t varies from 1 to 3. We estimate this for $i = 1 \ldots n$ officers, $t = 1 \ldots t$ time periods, and $\alpha_{i,t}$ represents case-specific effects including suspected crime and probable cause basis of past month stops. The predictors are present covariates of stops and lag the past months' hit rates to estimate learning and updating effects. Standard errors are clustered by officer.

Models were estimated in iterations where additional blocks of predictors were added: a baseline model accounting for the lagged number of hits per month; a model with officer-related variables, including length of career, command assignment, and prior months' stops, frisks, and searches; and a final model with parameters related to offenses and suspects, including measures of suspicion, reasons for the stops, and demographic information about those stopped.

The models also include a measure of the context (borough, in this case) where the weights are assigned to productive and unproductive bad stops. Those weights or values can be internalized by officers based on their experience and exposure.[153] Additionally, officers are active across precinct boundaries during their careers. Over the course of the study period, only about 30% of officers made stops entirely within one precinct, and it was common to observe an officer making a stop in different precincts within the same month.[154] Movement between boroughs occurred, but was less prevalent, with over 60% of officers making stops entirely within one borough.[155] Accordingly, borough may provide a more accurate context in which officers observe and weigh signals of suspicion than the individual precincts. In this vein, the variables for whether an officer was predominantly assigned to a housing, transit, or patrol unit provide insight into whether

[153] *See, e.g.*, *Bureaus*, NYPD, https://www1.nyc.gov/site/nypd/bureaus/bureaus.page [https://perma.cc/ETY2-FW2G]. We incorporate borough, rather than precinct. Although most stops are made by officers assigned to precincts, several NYPD commands assigned to intensive enforcement details are boroughwide, including Anti-Crime and Special Narcotics, and units focused on gang activity. Similarly, management of NYPD patrols are decentralized to borough commands; crime rates also vary considerably by borough. *See* PETER ZIMROTH, NINTH REPORT OF THE INDEPENDENT MONITOR 9 (2019) [hereinafter ZIMROTH, NINTH REPORT], https://ccrjustice.org/sites/default/files/attach/2019/01/Monitor.pdf [https://perma.cc/ZLC3-GQXZ].

[154] Across the nearly 50,000 officers in the study, only about 16,600 made stops only in one precinct. This may be a result of officers moving commands or positions, or officers working at a boroughwide command.

[155] In order to examine the behavior of those who did move between boroughs, we operationalize six borough categories: Brooklyn, Bronx, Manhattan, Queens, and Staten Island, and a sixth "borough" for those who moved between boroughs during their career.

learning occurs differentially based on a boroughwide versus precinct-based experience.

For the Bayesian analysis, the count of current month hits is again the dependent variable with the same one-to-three-month lag structure. We utilize default, uninformative, normal priors. The model assumes that the "hit rate" (seizure, arrest, or any other measure of productive stops) in a time period will be a function of the number of productive and unproductive stops in prior periods, and the features of those stops. Features should include characteristics of the person(s) stopped in prior stops, the suspected crime(s) in those stops, and the suspicion indicated for those stops. Because race is a central feature of the *Floyd* litigation and the patterns of stops during that time, we include parameters for whether the suspect was Black or Latinx. Additionally, as with the previous models, we include a parameter for the suspicion basis of the stops. We also apply the same iterative modeling process as with the frequentist regressions.

The Bayesian analysis assumes that over a series of decisions, officers will incorporate factors that were part of an earlier successful event, and discard factors from earlier negative events, and improve each decision by adjusting or recalculating their prior estimates of success and the features that produce them. Unlike standard economic theory, where the estimates of the factors that produce success or failure remain unchanged over time (and are an average of prior events), we assume that an officer's prior views of the salience of relevant features of the decision context change over time. So, we ask whether the outcome of an officer's decision to conduct stops produces new information that is incorporated into a reweighting of the features of the next potential encounter to revise their prior belief and form an updated, or posterior belief. In Bayesian terms, this is a process of updating prior beliefs.[156] In a dynamic sequence of decisions, an officer should put a positive weight on the signal or combination of factors that produced a positive outcome, thus changing her perceptions of the information or signal available for the next event. Ideally, failure should also signal information to update prior beliefs. But since police face no cost for a failed stop, they are likely to either ignore or downweigh the signal from that event.

The formal expression of the model is:

[156] Shamena Anwar & Thomas A. Loughran, *Testing a Bayesian Learning Theory of Deterrence Among Serious Juvenile Offenders*, 49 CRIMINOLOGY 667, 670 n.5 (2011) (stating that "Bayes' theorem provides a formula for the conditional probability some event A occurs given that another event B has already occurred, which can be written as $P(A \mid B) = P(A \cap B) / P(B)$); Benjamin R. Baer & Martin T. Wells, *A Bayesian Approach to Event Studies for Securities Litigation*, 176 J. INST. & THEORETICAL ECON. 115, 116–18 (2019).

$$\theta_{it} = \delta_{it}\,(A_{it}/C_{it}) + (1 - \delta_{it})s_{it,}$$

where A and C are parameters of "good" and "bad" stops, and *s* is a parameter for contextual factors. As information is incorporated from experience, the officer will form a new posterior perception that is a weighted average of the two types of stops:

$$p_{i,t} = \alpha_{i,t}\,\theta_{i\,t} + (1-\alpha_{i,t}\,)\,p_{i,t-1}$$

where αi,t ∈ (0,1) denotes the relative weight on the signal for individual *i* in period *t* .

We then combine the two equations to form the dynamic updating process:

$$p_{it} = \alpha_{i,t}\,\delta_{it}\,(A_{it}/\,C_{it}) + (1-\alpha_{it}\,)\,\mathrm{p}_{i,t-1} + \alpha_{it}\,(1-\delta_{it}\,)\,\mathrm{s}_{it}$$

We use a similar iteration of predictors in the Bayesian analysis as in the frequentist learning model.

III. Results

A. Officer-Level Descriptive Analysis

Figures 1–5 show descriptive trends for all stops, across officers for the entire 2004–2016 period. The left axis in Figure 1 shows that stops rose and remained high from the start of the study period until class certification was granted in the *Floyd* litigation in May 2012, then fell precipitously, with an accompanying change in the rate of recovery of weapons.[157] The number of stops peaked in 2011 at approximately 686,000, and by 2013 was lower than 200,000 and continued to fall to about 12,000 in 2016.

The number of individual officers making one or more stops each year is shown on the right axis of Figure 1.[158] The number of unique officers making at least one stop that year declined slowly from 23,000 in 2005 to about 18,000 in 2012. The number then fell to fewer than 15,000 in 2013, and then to fewer than 5,000 in 2016.[159] To test for the effects of this attrition from stop activity on the model estimation, we re-estimated all models limiting the time window to stops through 2013. The results were unchanged.

[157] Mummolo, *supra* note 10, at 12.

[158] Each of these individual officers may have made one or more stops in prior or subsequent years and is counted in each year the officer made a stop.

[159] There is some question as to the reporting rates starting in 2013. *See* Ryan Devereaux, *NYPD Stop-and-Frisk Memo Revealed in Civil Rights Court Battle*, GUARDIAN (Mar. 27, 2013), https://www.theguardian.com/world/2013/mar/27/nypd-stop-and-frisk-memo [https://perma.cc/7NJC-85UE]. Rates of reporting per stop are estimated now at less than 40%. ZIMROTH, NINTH REPORT, *supra* note 153, at 40–41. As a sensitivity check, the authors also ran the Poisson models only through 2013; the results were consistent with the models through 2016.

Additionally, a differential number of individual officers first appear in the data each year (*e.g.*, first make a stop). In 2004, the first year available, the data contain nearly 20,000 unique officers, which likely includes officers whose first stops were in 2004 and those whose first stops were prior to 2004. In 2005, an additional 10,000 joined the dataset. Between 2013 and 2016, roughly 1,000 new officers appeared in the data each year. In total, this accounts for the 49,000 unique officers in the dataset across 2004 through 2016. Narrowing the sample to those officers who first started between 2006 and 2015, there are 18,000 unique officers, a much smaller number, as fewer officers joined the dataset over time.

Obviously, stop activity is not distributed equally across all officers. Across the study period for all 49,000 officers, the median number of stops per officer who made one or more stops was thirty-one, for all officers regardless of length of time in the panel, while the average was 100 stops. The lowest 25th percentile of officers made only three stops over the 13-year period, while the 75th percentile made 127 stops over that same period. Approximately 8,000 officers made only one stop over the course of the study period, while nearly 1,800 made 500 or more stops. The wide variation in stop frequency may reflect the fact that not all officers were present throughout the entire study period, censoring their stop activity based on their entry and exit from the cohort of officers making stops. Some were present for only a few months, while others were present for years, and we control for career length in the estimates of updating. On average, officers appeared for 3.7 years, not necessarily consecutively, and about 28% of officers were present for less than one year. This may indicate not only that some officers may have left the NYPD, but also that some officers moved between commands with differing stop responsibilities and exposures or that some officers simply made fewer stops, since officers only appear in the dataset if they made a stop.

Figure 1. Total Stops and Officers Making Stops per Year

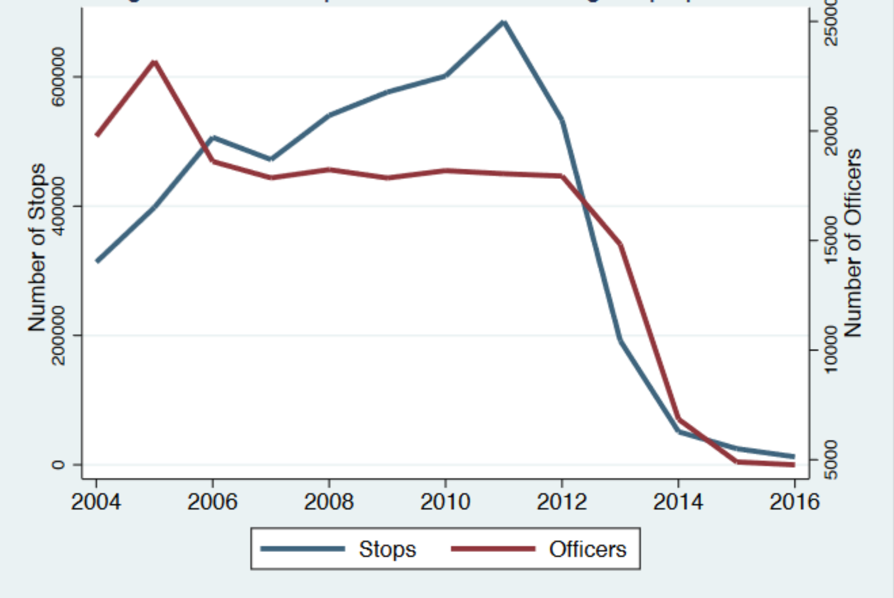


Figure 2 shows results for the various measures of hits over time. Each hit rate remained below 15% of all stops through 2013. After that, the hit rate for arrests and summonses rose to nearly 25% following the *Floyd* trial, but hit rates for seizures of all types rose slowly after 2013. Substantively, hit rates were consistently lower for weapons or contraband recovery across the panel years, remaining below 10%, despite small increases following the *Floyd* litigation. [160]

[160] Calculated based on averaging each officer's yearly hit rate across all officers making stops that year.

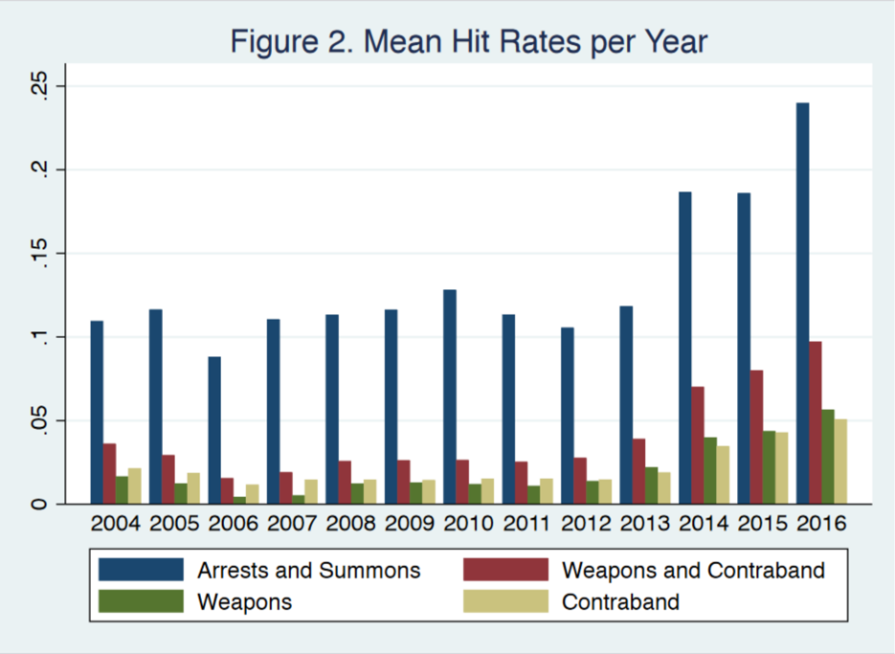


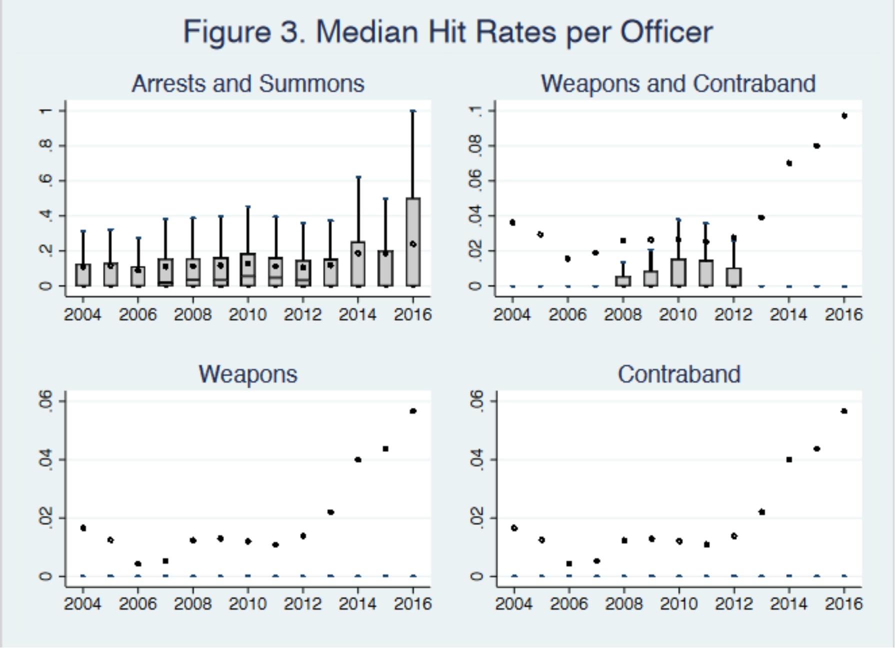
Figure 2. Mean Hit Rates per Year

Figure 3 shows median hit rates, given that the mean hit rates are substantially affected by officers at the margins whose hit rates were particularly high. In contrast to the increase in hit rates after 2013, the median hit rates remained consistently low throughout the study period. Figure 3 plots histograms for each hit rate measure overlaid by the median hit rates per year for that measure. Whether the increase in average hit rates after the *Floyd* verdict in 2013 represents a sudden, and admittedly hard to explain, uptick in updating and learning is not obvious. It may represent more careful decision-making about stops in response to agency mandates and the scrutiny of a federal court monitor on stop activity.[161] A complementary explanation is that testimony in the 2013 *Floyd* order marked the revelation of quotas for stops imposed by police executives, and their undoing pursuant to the court's oversight and removal.[162] Regardless of the reasons, the number of officers

[161] *See* Floyd v. City of New York, 283 F.R.D. 153, 162–69 (S.D.N.Y. 2012) (describing evidence of stop quotas and constitutional violations stemming from NYPD's stop and frisk program); *see generally* Memorandum from James Hall, *supra* note 15 (standardizing entries officers had to make when submitting a UF250 report).

[162] Margot Adler, *At 'Stop-And-Frisk' Trial, Cops Describe Quota-Driven NYPD*, NPR (Mar. 21, 2013), https://www.npr.org/2013/03/21/174941454/at-stop-and-frisk-trial-cops-describe-quota-driven-nypd [https://perma.cc/H2N6-7RWU].

making stops decreased, as did their stop activity, while the returns from those stops increased as officers applied discretion to their exercise of the stop authority, rather than responding to internal mandates that were orthogonal to the requirements of both state and federal law. [163]



The median arrests and summonses hit rate peaked in 2010, at 5.6%, fell to near zero by 2013, and continued at zero through the end of the study period. However, at the margins, there was improvement. At the 75th percentile, the hit rate was 12.5% in 2004. By 2016, it had reached 50%. The overall improvement in hit rates for weapons after 2013 coincided with the reduction in stops, suggesting that the reduced pressure of quotas and institutional pressures contributed to more careful selection of suspects for stops and (in some cases) searches.[164] The officers at the margins, as the study period progressed, became less like their counterparts at the median—at least considering arrests and summonses.

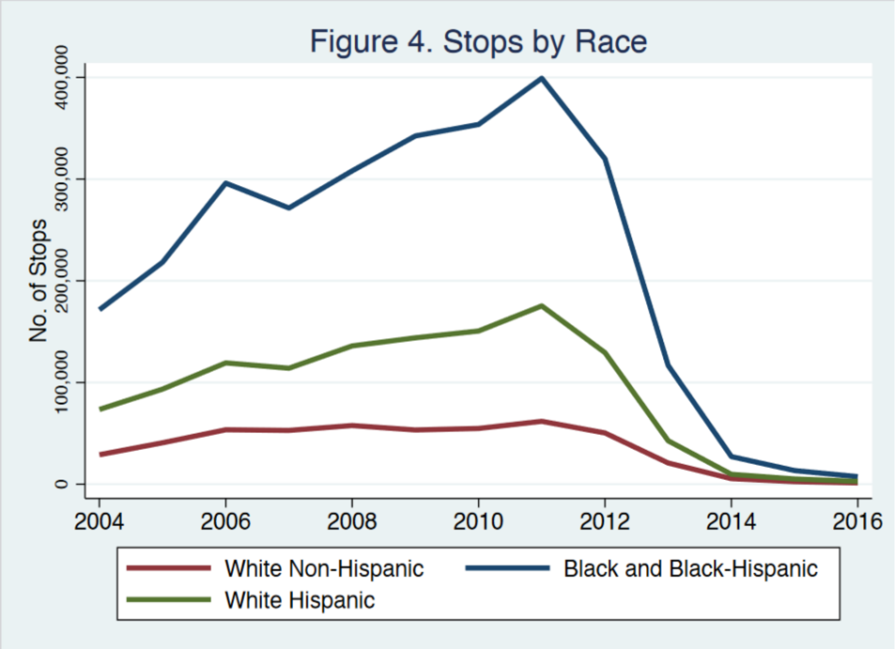
However, this substantial improvement at the margins is not replicated when considering the hit rates for weapons and contraband (both combined

[163] Calculated based on each officer's hit rates per year, aggregated over all officers making stops that year.

[164] Adler, *supra* note 162.

and separated). Throughout the study period, the median hit rates for these measures were near zero, and, in nearly every year, the 75th percentile was also near zero. For the combined rate, the only time the 75th percentile hit rate was not zero percent, it rose only from 0.5% in 2008 to 1% in 2012 before falling to zero the following year. The low hit rates for weapons and contraband appear to be resistant to improvement at the median, even as stop rates fell and the *Floyd* litigation brought increased scrutiny to officers' stop practices. And while our analysis focuses on rates specific to officers, other studies report no changes overall from 2012 to 2015 in the rates of recovery of firearms or contraband by suspect race.[165]

Figure 4. Stops by Race



The stop and hit rates varied by suspect race and ethnicity. Stop rates were higher for non-whites compared to whites (Figure 4) throughout the 13-year period. Although stop rates fell dramatically for all groups after the

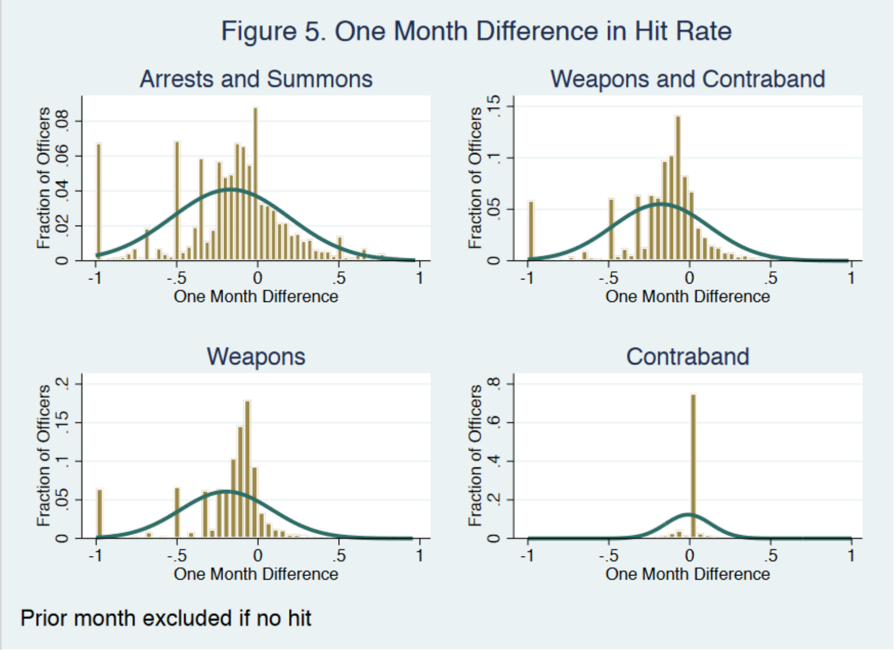
[165] John MacDonald & Anthony A. Braga, *Did Post*-Floyd et al. *Reforms Reduce Racial Disparities in NYPD Stop, Question, and Frisk Practices? An Exploratory Analysis Using External and Internal Benchmarks*, 36 JUST. Q. 954, 973 (2019) (showing random variation by year in the odds of recovery of a weapon pursuant to a stop for Black and Latinx suspects relative to whites and others). In each instance, the annual odds ratio was lower for each racial and ethnic group. compared to the recovery rates for whites. *Id.*

Floyd order in 2013, racial disparities remained.[166] This was the basis for the *Floyd* court's characterization of police practice as "indirect profiling," where officers make significantly more stops of non-white suspects with significantly lower hit rates overall. There were small differences in hit rates by race, but those were observed based on low percentages of arrests or seizures over a large denominator of stops. Why officers were more accurate for white suspects than non-white suspects and why their learning coefficients suggest a decline over time are questions addressed in the updating analyses.

Finally, Figure 5 shows the unadjusted hit rate differences from one month to the next for each of the outcome measures, where the officer had at least one successful stop in the prior month. For arrests and summonses, the hit rates in the month after a successful hit declined for most officers. For seizures, there was a slight decline in hit rates in the following months overall, with large decreases for a small number of officers. But there also were small increases for a small number of officers, offsetting the decreases for others. For contraband, this was particularly evident, with a large group of officers showing small increases in their hit rates. These distributions suggest considerable heterogeneity in learning and updating over the short time periods.[167] We test these increases and declines with a series of models that introduce controls for the context of the stops and for longer periods of observation.

[166] ZIMROTH, FIFTH REPORT, *supra* note 66, at 46 ("By 2015, the correlation (rho) between the amount of crime in these small areas relative to the number of stops in the same areas had diminished substantially . . . ."). Given the low stop rates, the estimation of racial disparities in the stop rates after 2013 were sensitive to model specification and measurement decisions.

[167] Sidney J. Winter, *The Satisficing Principle in Capability Learning*, 21 STRATEGIC MGMT. J. 981, 987, 991 (2000) (showing heterogeneity in learning ability and adjustment speed in response to a "wake-up call").



B. UPDATING AS A LINEAR PROCESS

We estimated random effects Poisson panel regressions to examine whether police officers' hit rates improve from one month to the next, and whether learning unfolds over longer periods between stop events. For each hit rate measure, we first estimated hit rates for simple models with prior month hit rates for the three months preceding the observation month, and controls for the borough of the stop. Later iterations included frisk and search rates in the prior months, a measure of behavioral indicia of suspicion (probable cause stops), officer command, and total stop activity. Additional controls were then added in the final iteration for the proportion of minority individuals, those aged between sixteen and twenty-four, and males stopped in prior months, as well as the proportion of stops in prior months based on suspected crime types. The exposure variable was the total number of stops per officer for each period, producing estimates of hit rates that are conditional on stop activity. Both these and the Bayesian regressions examine those officers whose first stop was between 2006 and 2015. As a sensitivity check, we also ran the Poisson models only through 2013 to see if the post-*Floyd* era created a selection effect on the sample of officers making stops; the results were consistent with the models through 2016.

Table 1 shows the results of the regression model for the first of four hit rate measures: arrests made and summonses issued.[168] Across all model specifications, past months' hit rates are consistently small but significant and positive predictors of the current month's hit rate, suggesting evidence of continuity, if not learning, in both arrests and seizures. In models 2 and 3, which include the additional controls, the positive coefficient associated with the one-month lagged hit variable indicates a 10% increase in hits compared to the prior month. We also find that prior hit rates in the preceding three months predict both arrest and seizure rates in the following month.

Officers whose stop careers began after the *Floyd* litigation break (2012) had higher success rates compared to officers whose careers unfolded during the pre-*Floyd* years of peak stop activity, although these differences were only weakly significant. Overall, stops after the *Floyd* verdict were less likely to result in either arrests or recovery of weapons; stops after the *Floyd* trial break demonstrate an almost 10% decline in hits compared to those stops before the break. The differences in officer hit rates by career start and year suggest a potentially important break in police norms and cultures that emerged in the recruits joining the police after the court's intervention. The connection between different officers and different stop activity suggests that there may be cohort effects that could explain contradictory stop outcomes and updating.

Stop legality affects learning over time. Although officers seem to gain knowledge on how to accurately form suspicion from past months' hits, those gains seem to dissipate as the suspicion criteria shift towards the narrower set of probable cause stops. These lower hit rates for probable stops across three time lags suggest the limits or counter-productive effects of a higher stop standard on hit rates that shift from appearances to observable behaviors.[169] Other studies suggest that probable cause stops may contribute to crime control,[170] an interesting contradiction between seizures and any deterrent effects of stops in the aggregate regardless of their outcomes.

[168] The Online Appendix is hosted on Digital Commons and can be accessed at https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?filename=0&article=4988&context=faculty_scholarship&type=additional [https://perma.cc/3QFT-7G7N] or https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?filename=0&article=7745&context=jclc&type=additional [https://perma.cc/FY8Y-EUZ2]. It includes full model results, including parameter estimates for suspected crime, suspect demographics, and officer patrol assignment (traffic, housing, transit, patrol).

[169] Samaha, *supra* note 151, at 1575 ("The general thought is that relatively accessible appearances sometimes . . . help make for a reality of interest."); Sampson, *When Things Aren't*, *supra* note 151, at 106 ("[O]ur perceptions of disorder and the consequences of acting on it are fundamentally social in nature rather than fixed in meaning.").

[170] Fagan, *supra* note 25, at 64 n.135; MacDonald et al., *supra* note 69, at 2–3.

Table 2 shows the results of Poisson regressions on seizures of weapons and contraband. The patterns are similar. Hits for all three prior lags are significant and positive across models, with roughly similar effect sizes to the arrests and summonses model. Officers who make more stops in the current period demonstrate a marginal, but significant decline in hits, suggesting that "less is more" when implementing a "program"[171] of stops instead of exercising "reasonable," "articulable" and "individualized" suspicions.[172] This marginal but significant decline holds for the number of stops one month prior, but dissipates for longer lags. Unlike the results in Table 1, the results here suggest hit rates that are neither significantly lower nor higher for stops taking place after the 2012 *Floyd* break; as with the results in Table 1, hit rates for weapons and contraband are only marginally but still significantly higher for those officers who first appear after the *Floyd* break.

[171] Meares, *supra* note 26, at 162.

[172] Terry v. Ohio, 392 U.S. 1, 21 (1968); *id*. at 37 (Douglas, J., dissenting); People v. De Bour, 352 N.E.2d 562, 573 (N.Y. 1976).

Table 1. Random Effects Poisson Regression for "Hits" for Arrests or Summons Issued for Officer Career for Three Lag Periods, New York City, 2006–16 (IRR, SE, p)

	Model		
	1	2	3
Hit_Lag1	1.075***	1.102***	1.099***
	(0.004)	(0.005)	(0.005)
Hit_Lag2	1.042***	1.054***	1.052***
	(0.003)	(0.003)	(0.003)
Hit_Lag3	1.027***	1.037***	1.035***
	(0.003)	(0.003)	(0.003)
Year of Stop		1.028***	1.030***
		(0.004)	(0.004)
Total Month Tenure		1.005***	1.005***
		0.000	0.000
Stop After Floyd Class Cert		0.885***	0.885***
		(0.014)	(0.014)
Officer Start After Floyd Class C		1.176*	1.171*
		(0.085)	(0.083)
Number of Stops		0.983***	0.983***
		(0.001)	(0.001)
Number of Stops_Lag1		0.983***	0.984***
		(0.001)	(0.001)
Number of Stops_Lag2		0.992***	0.992***
		(0.001)	(0.001)
Number of Stops_Lag3		0.992***	0.992***
		(0.001)	(0.001)
Search Rate_Lag1		1.164***	1.152***
		(0.026)	(0.026)
Search Rate_Lag2		1.142***	1.135***
		(0.025)	(0.025)
Search Rate_Lag3		1.135***	1.130***
		(0.024)	(0.024)
Frisk Rate_Lag1		1.035*	1.040**
		(0.015)	(0.016)
Frisk Rate_Lag2		1.025	1.036*
		(0.015)	(0.016)
Frisk Rate_Lag3		1.017	1.035*
		(0.014)	(0.015)
Percent Probable Cause Stops_Lag1			0.930***
			(0.013)
Percent Probable Cause Stops_Lag2			0.974*
			(0.013)
Percent Probable Cause Stops_Lag3			0.986
			(0.013)
Constant	0.078***	0.000***	0.000***
	(0.002)	(0.000)	(0.000)
N	200535	200535	200535

Notes. Contraband includes drugs, stolen property, or other unauthorized possession of goods. Results shown as IRR. Models estimated with standard errors clustered by officer. Controls for year of stop, year of officer first stop, reason for stop, officer command, percent of stops of persons ages 16-24, percent of stops by suspect race or ethnicity, suspected crime, borough of stop, and *Floyd* litigation year.

Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2. Random Effects Poisson Regression of "Hits" for Weapons or Contraband Seizures over Officer Stop Career for Three Lag Periods, New York City, 2006-16 (IRR, SE, p)

	Model		
	1	2	3
Hit_Lag1	1.133***	1.118***	1.113***
	(0.019)	(0.017)	(0.016)
Hit_Lag2	1.083***	1.065***	1.060***
	(0.015)	(0.014)	(0.014)
Hit_Lag3	1.077***	1.058***	1.054***
	(0.014)	(0.012)	(0.012)
Year of Stop		1.114***	1.120***
		(0.009)	(0.009)
Total Month Tenure		1.004***	1.004***
		(0.001)	(0.001)
Stop After Floyd Class Cert		0.959	0.957
		(0.030)	(0.030)
Officer Start After Floyd Class C		1.358**	1.379**
		(0.144)	(0.144)
Number of Stops		0.973***	0.973***
		(0.002)	(0.002)
Number of Stops_Lag1		0.995***	0.996**
		(0.001)	(0.001)
Number of Stops_Lag2		1.000	1.000
		(0.001)	(0.001)
Number of Stops_Lag3		0.996**	0.997*
		(0.001)	(0.001)
Search Rate_Lag1		1.452***	1.420***
		(0.089)	(0.085)
Search Rate_Lag2		1.358***	1.334***
		(0.069)	(0.066)
Search Rate_Lag3		1.270***	1.254***
		(0.061)	(0.059)
Frisk Rate_Lag1		1.206***	1.236***
		(0.039)	(0.042)
Frisk Rate_Lag2		1.100**	1.116***
		(0.034)	(0.037)
Frisk Rate_Lag3		1.071*	1.107***
		(0.030)	(0.034)
Percent Probable Cause Stops_Lag1			0.973
			(0.025)
Percent Probable Cause Stops_Lag2			0.943*
			(0.025)
Percent Probable Cause Stops_Lag3			0.949*
			(0.025)
Constant	0.015***	0.000***	0.000***
	(0.001)	(0.000)	(0.000)
N	200535	200535	200535

Notes. Contraband includes drugs, stolen property, or other unauthorized possession of goods. Results shown as IRR. Models estimated with standard errors clustered by officer. Controls for year of stop, reason for stop, officer command, percent of stops of persons ages 16-24, percent of stops by suspect race or ethnicity, suspected crime, borough of stop, and *Floyd* litigation year.

Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In both Tables 1 and 2, the lagged search rates are significant predictors of subsequent hits, in addition to prior months' hits. The rates for arrests and summonses hits range from a factor of 1.15 (one-month lag) to 1.13 (two- and three-month lags). The larger effect sizes for weapons and contraband hits from frisks and searches in prior stops appear to be largely driven by the returns seen on hits for contraband only, not on weapons, as indicated in Appendix Table 1. That is, while the lagged search rate is only weakly significant for subsequent seizures of weapons, those rates are highly significant for seizures of other contraband. For instance, an increase in the prior month's search rate is associated with an increase in the hit rate for contraband of about a factor of 1.6. Frisk rates tell a slightly different story. While the prior frisk rate is a significant positive predictor for subsequent seizures of weapons and contraband, for the arrests and summonses measure, the lagged frisk rates, even only lagged by one month, are only marginally significant. The results point to a learning process stemming from prior experience searching suspects, with prior frisk experience playing a more minor role in the process.

These tables, as well as the full models in Appendix Table 1, indicate the coefficients associated with month tenure are positive and highly significant for the measures of arrests and summonses; weapons and contraband; and contraband. However, the effect sizes are very close to 1.0, which suggests little difference in incidence of successful stops, given an increased number of months making stops.

Table 3. Random Effects Poisson Regressions of Stops on Four Outcomes for Three Lag Periods, New York City, 2006-16 (IRR, SE, p)

	Stop Outcome			
	Arrests/ Summons	Weapons/ Contraband	Weapons	Contraband
Hits_Lag1	1.099^{***}	1.113^{***}	1.100^{**}	1.136^{***}
	(0.005)	(0.016)	(0.037)	(0.022)
Hits_Lag2	1.052^{***}	1.060^{***}	1.088^{**}	1.055^{**}
	(0.003)	(0.014)	(0.032)	(0.018)
Hits_Lag3	1.035^{***}	1.054^{***}	1.051^{*}	1.061^{***}
	(0.003)	(0.012)	(0.025)	(0.017)
Constant	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
N	200535	200535	200535	200535

Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 summarizes results for the third model in Tables 1 and 2, controlling for full covariates, as well as for the updating models for weapons and contraband separately. The results show the consistent learning that takes place among those officers who either make arrests or seize weapons or contraband in any given month. Not surprisingly, effect sizes decrease as the lags increase, with the one-month lag variable associated with the largest hit rate ratio increase.

However, recall from Figures 1–3 that overall, the large majority of officers have either no or very low stop and hit activity, tempering the practical significance of the results. While there is evidence of learning, its effect is likely concentrated among those officers with higher levels of positive stop outcomes overall, in particular, because increases in stop activity seem to have a negligible (and negative) effect on hits. Although seizure rates improved over the 13-year study period, they still remained low over time, never rising above 5% until 2014, and the marginal contribution of these factors to improved seizure rates were low but significant.[173] Given the low baselines of hits as well as the small effect sizes in some instances, it is not unreasonable to temper the findings by noting their limited practical significance.[174]

In the full models shown in Appendix Table 1, the negative, largely insignificant coefficients for Black and Latinx suspects are consistent with the broader racial currents that animated the controversies over *Terry* stops in New York. Some lags show significant negative effects, others show no effects. For stops resulting in arrests or summonses, the percent of Black or Latinx suspects stopped in prior months has no significant effect on the current month's hit rate, and for seizures of weapons and contraband, there is a weakly significant decreasing rate ratio. There are negative effects for the two-month lag, but only for that period. The coefficients for Black and

[173] Some courts have argued for a test based on the efficacy of stops in detecting crime or locating contraband, but here too, there is no agreement on what constitutes an acceptable "hit rate" that satisfies the reasonableness standard. In his dissent from *Navarette v. California*, 572 U.S. 393, 410 (2014), for example, Justice Scalia suggested that at least five, if not ten percent, of the entire universe of incidents would need to be an accurate "hit" to be indicative of reasonable suspicion. According to Scalia, absent such a showing, the basis of suspicion is not reasonable. *Id.* Hit rates in this study for weapons or contraband did not rise above 5% overall until well after the *Floyd* opinion and order were issued in 2013.

[174] *See, e.g.*, Michael J. Peeters, Practical Significance: Moving Beyond Statistical Significance, 8 CURRENTS PHARMACY TEACHING & LEARNING 83, 84 (2016); *see also* Stanley Pogrow, *How Effect Size (Practical Significance) Misleads Clinical Practice: The Case for Switching to Practical Benefit to Assess Applied Research Findings*, 73 AM. STATISTICIAN 223 (2019).

Latinx suspects suggest that any lessons of past successes in arrests or seizures do not carry forward for these suspects, who were the subject of nearly 85% of the stops through 2013. As with the earlier results, whatever learning may take place in the search for weapons seem to carry forward only for white suspects. The factors that were identified by Simoiu et al. in their inframarginality analysis that predict weapons seizures seem to not be implemented over time and through millions of police stops that were conducted from over the ten-year study period.[175] Accordingly, we might conclude that the learning that took place was not randomly distributed across the full range of civilians stopped and searched by the police, but for only a non-random and primarily white subset of those stopped.

Consistent with the *Floyd* court interpretation of "indirect profiling," the results suggest that Black and Latinx suspects are, on average, no more or less likely across three time lags to be sanctioned pursuant to a stop. But other research with these data suggest sanctions seem to be less of an intervention in serious crime than a processual punishment for minority suspects.[176] Again, the analyses of post-stop arrest outcomes suggested that these arrests tend to be either dismissed or pled to minor crimes, mostly quality-of-life crimes.[177] This suggests a pattern akin to the managerialism and social control that describes misdemeanor justice for those charged with the least serious crimes in urban policing and criminal courts.[178]

The regression analyses present complicated answers to the question of updating and learning by officers. First, we do find consistent evidence of updating and learning using based on the lagged hit measure, but not necessarily for those stops based on the most carefully articulated indicia of

[175] *See* Simoiu et al., *supra* note 9, at 1198 ("Given that one cannot rule out the possibility of such signal distributions arising in real-world examples (and indeed we later show that such cases do occur in practice), the benchmark and outcome tests are at best partial indicators of discrimination.").

[176] Jeffrey Fagan & Elliott Ash, *New Policing, New Segregation: From Ferguson to New York*, 103 GEO. L.J. ONLINE 33, 52–55 (2017).

[177] SCHNEIDERMAN, *supra* note 77, at 1, 6, 11[77]. Quality of life offenses are defined by the NYPD as "aggressive panhandling, squeegee cleaners, street prostitution, 'boombox cars,' public drunkenness, reckless bicyclists, and graffiti." *See* NYPD, POLICE STRATEGY NO. 5: RECLAIMING THE PUBLIC SPACES OF NEW YORK 5 (1994), https://www.ncjrs.gov/pdffiles1/Photocopy/167807NCJRS.pdf [https://perma.cc/9YEF-4ZVW]. In 2015, NYPD has expanded its definition of quality-of-life policing as "enforcing a variety of laws against street drug dealing, public drinking, public marijuana smoking, open-air prostitution, and other minor offenses." *See* NYPD, TACKLING CRIME, DISORDER, AND FEAR: A NEW POLICING MODEL 2 (2015), http://www.nyc.gov/html/nypd/html/home/POA/pdf/Tackling_Crime.pdf [https://perma.cc/UAR9-6CMB].

[178] *See* ISSA KOHLER-HAUSMANN, MISDEMEANORLAND: CRIMINAL COURTS AND SOCIAL CONTROL IN THE AGE OF BROKEN WINDOWS POLICING 3–20 (2019).

suspicion or for increased stops more generally. This seems to contradict the assumptions of disciplined formation of suspicion based on probable cause, but the effect may reflect more about the institutional and programmatic context than about officer perceptions and judgments. In other words, and as noted earlier, the negative contemporaneous effect may have less to do with officer judgment than with the pressure to steadily increase stops over time, regardless of their legal basis.[179] The declining rate ratio coefficient for the lagged percentage of stops based on probable cause may reflect more about the institutional context—the mandate to increase stops—than about learning opportunities from making disciplined versus inchoate stops. Officers appear to negatively update from their actions in the past month when they apply narrower indicia of suspicion, where stops with more subjective bases of suspicion seem to positively affect current hit rates. This is not good news for the Fourth Amendment prong of the *Floyd* opinion, where the court cited both poor results of stops and sharp racial disparities in both the rationales for stops together with their poor hit rate.[180]

C. ARE POLICE OFFICERS BAYESIANS?

We estimated Bayes regressions that were structured similarly to the frequentist models discussed in the previous Section. Results were iterated from the same baseline model with only the past months' hits and borough controls, with additional blocks of predictors including legal features of the stops and covariates added to later models. Results of the final models are shown in Table 4 for each measure of stop success. Estimates for prior months' hits suggest learning from one month to the next, similar to the prior three-month lag parameter in the Poisson models. Here, the results indicate a 95% likelihood that the parameter for the current month's hits increases from about 14% (arrest or summons) to 23% (contraband) as the prior month's hits increase. There also are similar but slightly weaker results for the two and three-month lags. Appendix Tables 2–5 show that the results in each model are robust to the inclusion of blocks of legal and demographic covariates, and

[179] Fagan, *supra* note 25, at 55–56.

[180] Floyd v. City of New York, 959 F. Supp. 2d 540, 572–75 (S.D.N.Y. 2013); Fagan & Geller, *supra* note 85 at 51 ("The results suggest that the observed patterns of narratives have evolved into shared narratives or scripts of suspicion, and that these patterns are specific to suspect race and neighborhood factors."); Grunwald & Fagan, *supra* note 69, at 398 n.157 ("Other subjective factors may also include suspicious bulges, sights or sounds of crime, or evasive actions. These factors are likely vulnerable to cognitive distortion and bias, especially in the context of race or threatening situations." (citing Jennifer L. Eberhardt, Phillip Atiba Goff, Valerie J. Purdie & Paul G. Davies, *Seeing Black: Race, Crime, and Visual Processing*, 87 J. PERSONALITY & SOC. PSYCH. 876, 880 (2004)).

Table 4. Bayesian Poisson Regression for “Hits” for Contraband Seizures for Officer Career for Three Lag Periods, New York City, 2006–16 (IRR, 95% Credible Interval)

	Arrests/Summons			Weapons/Contraband			Weapons			Contraband		
	IRR	95% Cred. Interval		IRR	95% Cred. Interval		IRR	95% Cred. Interval		IRR	95% Cred. Interval	
Hit_Lag1	1.136	1.135	1.136	1.194	1.193	1.195	1.168	1.168	1.169	1.238	1.237	1.238
Hit_Lag2	1.073	1.073	1.073	1.131	1.129	1.131	1.173	1.172	1.174	1.119	1.119	1.120
Hit_Lag3	1.063	1.063	1.063	1.117	1.116	1.118	1.094	1.093	1.095	1.155	1.154	1.155
Year of Stop	1.028	1.028	1.028	1.087	1.087	1.087	1.122	1.122	1.122	1.082	1.082	1.082
Total Month Tenure	1.003	1.003	1.003	1.003	1.002	1.003	0.999	0.998	0.999	1.006	1.006	1.007
Stop After Floyd Class Cert	0.877	0.877	0.877	1.035	1.034	1.036	1.081	1.081	1.083	1.016	1.015	1.017
Officer Start After Floyd Class Cert	1.132	1.131	1.132	1.228	1.226	1.230	0.991	0.990	0.992	1.538	1.537	1.539
Number of Stops	0.985	0.984	0.985	0.978	0.977	0.979	0.975	0.974	0.976	0.982	0.981	0.982
Number of Stops_Lag1	0.979	0.979	0.979	0.996	0.995	0.997	0.998	0.998	0.999	1.000	0.999	1.001
Number of Stops_Lag2	0.990	0.990	0.991	1.000	1.000	1.001	1.000	0.999	1.001	1.006	1.005	1.007
Number of Stops_Lag3	0.989	0.989	0.989	0.999	0.998	1.000	1.001	1.001	1.002	0.999	0.999	1.000
Search Rate_Lag1	1.201	1.201	1.202	1.613	1.611	1.614	1.556	1.554	1.557	1.824	1.823	1.826
Search Rate_Lag2	1.178	1.177	1.178	1.424	1.421	1.426	1.419	1.418	1.421	1.492	1.490	1.493
Search Rate_Lag3	1.137	1.137	1.138	1.284	1.281	1.285	1.287	1.285	1.288	1.336	1.335	1.337
Frisk Rate_Lag1	1.061	1.060	1.061	1.317	1.316	1.318	1.249	1.248	1.250	1.373	1.372	1.374
Frisk Rate_Lag2	1.061	1.060	1.061	1.176	1.176	1.177	1.083	1.083	1.084	1.264	1.263	1.265
Frisk Rate_Lag3	1.059	1.059	1.060	1.180	1.178	1.181	1.137	1.135	1.137	1.205	1.204	1.206
Percent Probable Cause Stops_Lag1	0.942	0.942	0.943	0.970	0.969	0.972	0.882	0.881	0.882	1.028	1.027	1.030
Percent Probable Cause Stops_Lag2	0.990	0.990	0.990	0.928	0.927	0.929	0.878	0.878	0.879	0.970	0.970	0.971
Percent Probable Cause Stops_Lag3	0.995	0.994	0.995	0.943	0.942	0.944	0.917	0.916	0.918	0.936	0.935	0.937
Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N	200535			200535			200535			200535		

Notes. Random-walk Metropolis-Hastings sampling. Contraband includes drugs, stolen property, or other unauthorized property possession. Non-informative, default priors are used to estimate to variation in the number of stops in the current month. The results are shown in Table 4.

The Bayesian results are largely confirmatory of the frequentist estimates of the legal and demographic features of the stops. Hits in the current month lag are negatively associated with an increase in prior months' stops based on the more demanding probable cause level of suspicion in all measures besides the one-month lagged contraband only hits, although that parameter is relatively close to an incidence rate of 1.0. Hit rates are higher, controlling for the past month's successes and failures, when there are higher rates of frisks and searches in the three lagged months, although the effect sizes remain low for the arrests and summonses measure. This provides further support that the lessons gleaned from the rare successes in the past month carry forward to the next month. Having succeeded in a prior stop, officers seem to apply past knowledge of successes to conduct frisks and searches that are predicates of arrests or seizures, but the reasoning applied to make probable cause stops may not be as susceptible to learning based on exposure and experience.

Appendix Tables 2–5 shed some light on the effects of race. As with the Poisson models, the Bayesian estimates indicate that increased prior stops of Black and Latinx suspects likely to lead to fewer successful stops in the current month, despite the fact that the majority of suspects stopped are Black or Latinx. For arrests and summonses (Appendix Table 2), the parameter estimate is very close to 1.0, but for the measure of weapons and contraband seizures (Appendix Table 3), there is a 95% likelihood of a decrease between 10 to 20% in hits in the current month as officers stop more Black or Latinx suspects in prior months.

The effect of the *Floyd* litigation in Table 4 is, again, murky. For arrests and summonses, stops after the *Floyd* class certification order are associated with a little over a one-tenth decline in hits. For weapons or contraband hits, stops after the *Floyd* break are associated with a small increase in hits. Officers who first appear after the break are much more likely to have hits than those who started before, except for the measure of weapons only, where the rate ratio is nearly 1.0.

Officers seem to be able to apply at increasing rates accurate assessments of exactly which "crime is afoot," as demanded in *Terry* and subsequent cases,[181] although that remains a minority of all officers making successful stops. In fact, they seem to be particularly inartful in determining exactly what type of crime is "afoot," as evidenced by the suspicion parameter. This is particularly evident in their somewhat promiscuous use of the "high crime area" factor to form reasonable suspicion.[182] Others have analyzed the same data to identify a set of circumstances that predict that a suspect may be carrying a weapon, and perhaps these empirically derived factors are consistent with the intuitions identified here that suggest learning and updating.[183]

Evidence of learning and updating is situated in the longer arc of each officer's career. How well do officers perform in conducting stops over the course of their stop "careers"? Although there is evidence of learning and updating by officers, the learning is incremental over a very low base rate of arrests, summonses, and seizures of weapons and contraband.[184] The effect sizes reach significant levels, suggesting an increase of 5 to 20% in success rates per stop based on the immediate past outcomes, but they are over a very small year-to-year success rate.

[181] Terry v. Ohio, 392 U.S. 1, 30–31 (1968); *see also* Ornelas v. United States, 517 U.S. 690, 695–96 (1996) (discussing levels of suspicion required to "for suspecting the person stopped of criminal activity"); Illinois v. Gates, 462 U.S. 213, 235 (1983) (requiring a probability of criminal activity, rather than a prima facie showing).

[182] *See Floyd*, 959 F. Supp. 2d at 559; Fagan & Geller, *supra* note 85, at 70; Grunwald & Fagan, *supra* note 69, at 347.

[183] Goel et al., *supra* note 11, at 211–20.

[184] *See* ZIMROTH, FIFTH REPORT, *supra* note 66, at 11–12 tbl. 3, 40 tbl. 7; *Floyd*, 959 F. Supp. 2d at 558–59 (finding that between 2004 and 2012, only 1.5% of frisks resulted in finding a weapon, 6% of stops resulted in an arrest and 6% of stops resulted in a summons).

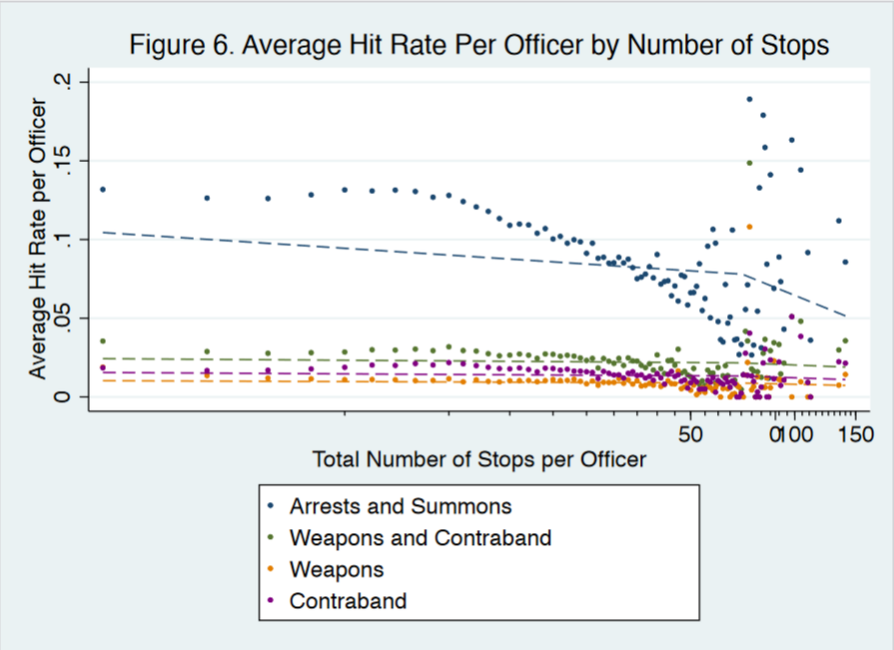


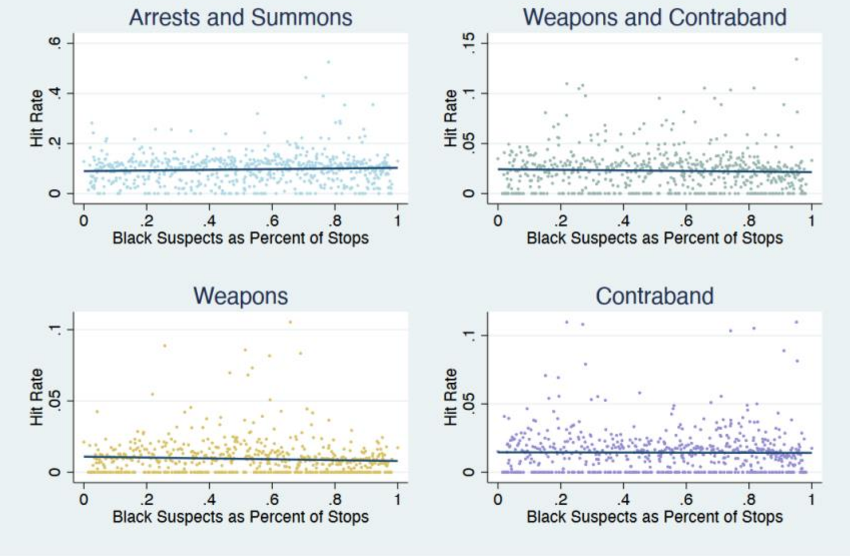
Figure 6. Average Hit Rate Per Officer by Number of Stops

Figure 6 illustrates why this might be the case. It shows the distribution of hit rates by the total number of stops for each of the outcomes measured: arrests and summonses; weapons and contraband; and weapons and contraband separately. In each case, despite evidence of learning in the regression models, officers' hit rates appear to decline, or at best stay relatively constant, as their level of stop activity increases. For arrests and summonses, there is a steady decline in hit rates as officer activity increases. The regression line in Figure 6 shows a decline from a hit rate of 10% for officers making fewer than twenty-five stops across their careers to less than 5% for officers making the highest numbers of stops. In fact, the regression line is pushed slightly upward by a small number of outliers at the upper end of the distribution. For weapons and contraband, hit rates appear insensitive to the total number of stops in an officer's stop career. The average hit rate declines slightly from approximately 2.5% overall at the low end of the distribution of stops across an officer's career to about 1.5% for officers at the upper end of the stop distribution. Officers who are successful tend toward continued success, but those officers who simply increase their number of stops do not seem to be learning from those unsuccessful ones.

Finally, Appendix Tables 1–5 demonstrate that race may complicate officers' learning and updating. As officers are exposed to more Black and

Latinx suspects,[185] their stop outcomes do not increase in accuracy, and in some instances, actually decrease their accuracy. These findings are illustrated in Figures 7 and 8, where increased stops of Black or Latinx suspects result in hit rates that are the same regardless of officers' rate of stops of non-white suspects. Given that stops of Black and Latinx individuals historically have made up the bulk of those stopped, this is concerning. Perhaps this suggests a lack of training on or internalization of the impact of race on suspicion. Moreover, our data lacked demographic information about the officers—such data would be helpful to better understand if an officer's own race influences their accuracy in stops, particularly for stops of Black and Latinx suspects.[186]

Figure 7. Average Hit Rates by Race (Black)



[185] Officers are exposed to more Black and Latinx suspects both because the allocation of officers to neighborhoods with higher concentrations of those populations, but also because of the skewing of perceptions of suspicion toward persons in each of those two groups. Jeffrey A. Fagan, *No Runs, Few Hits and Many Errors: Street Stops, Bias and Proactive Policing*, 68 UCLA L. REV. 1584, 1626, 1632–39 (2022) (showing disproportionate allocation of officers to predominantly non-white neighborhoods, and disparate treatment of persons by police once assigned to those neighborhoods).

[186] *See* Antonovics & Knight, *supra* note 10, at 169; *see also* Close & Mason, *supra* note 104, at 315–16 (showing that the probability that an African American male driver is searched is 0.43 percent when the stopping officer is African American, but it is 2.09 and 1.34 percent when the stopping officers are white male and Latino, respectively).

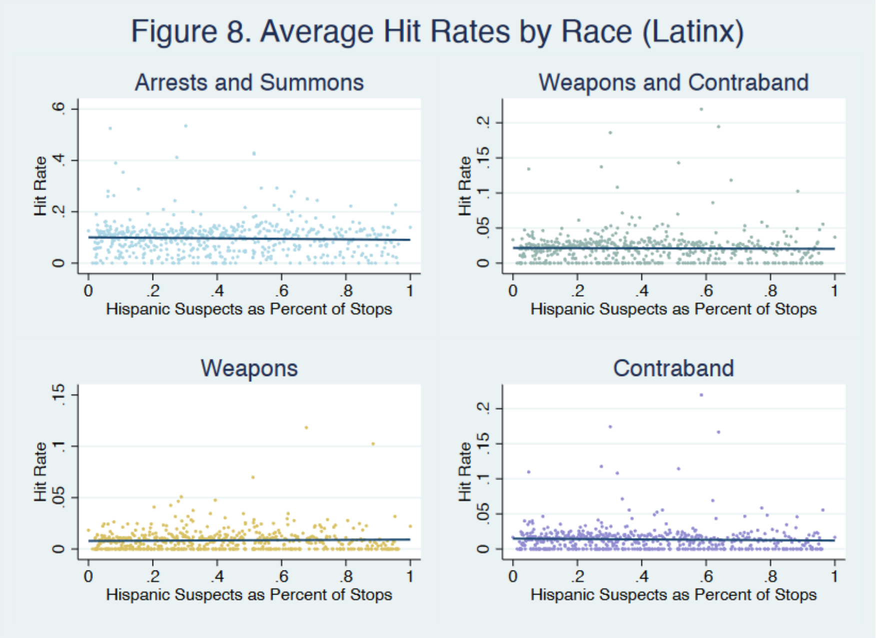


Figure 8. Average Hit Rates by Race (Latinx)

The combination of the results gives some credence to Taslitz's assertion that "[p]olice may suffer from inappropriately extrapolating from past results when they have insufficient information to identify a trend or an important factual distinction."[187] As shown, at the median, officers make very few stops per *year*; on a per month basis, officers are rarely making stops: if they are generalizing, they are generalizing from a very small sample of stops. And it seems that even when officers are proportionately more exposed to suspects of color, their hit rates do not improve.

IV. Discussion

Terry makes officer experience central to the constitutionality of a stop. It allows officers to "draw from the facts in light of [their] experience" and on specialized training to make inferences from and deductions about situations.[188] At the same time, the *Terry* Court warned against officers acting on "hunches," and demanded instead that they form a level of suspicion based on their own experience and the experience of others. From this language, Justice Douglas offered the term "reasonable suspicion." Left unsaid is the

[187] Taslitz, *supra* note 110, at 44.

[188] Terry v. Ohio, 392 U.S. 1, 27 (1968).

meaning of "reasonable." Is the formation of suspicion reasonable to the experienced officer's own prior judgments? The average judgments of other officers? And what is "reasonable" to the inexperienced younger officer? Perhaps she learns from others as her own experience accumulates.

Whether an officer makes a stop based on her own prior experience, as Officer McFadden did, or based on the collective experience of others, we know little about how officers perceive cues of suspicion, and whether and how those cues and indicia are internalized and then applied to particular situations. Since these may be split second judgments, or heavily contextualized judgments, we know little of this complexity and how officers understand their own decision instincts or develop and accumulate expertise. One perspective is that officers apply their training using decision heuristics that are individualized to their backgrounds and experiences, and hopefully shaped by training and feedback. The workplace culture and norms are part of that decision background, as well. Whether officers are maximizers (seeking to maximize their returns) or satisficers (accepting "good enough" returns on their investments of time or resources) may explain much about their capacity and motivation to learn and apply the results of prior stop experiences.[189]

This Article aimed to shed some light on this process by examining the stop careers of NYPD officers. The descriptive results indicate that most individual officers make relatively few stops, that very few of the stops made result in a legal sanction, and that even fewer turn up contraband or weapons. Even during the height of stop and frisk, in 2011, the average number of stops per officer was thirty-eight per year. By 2016, the average was fewer than three. The fact that individual officers were making so few stops, but aggregate stop and frisk actions had such an enormous effect on the polity of New York City,[190] lends support to Meares' conclusion that stop and frisk is a program "carried out by a police force *en masse*," rather than an individualized practice.[191]

The results also show that while most officers were inaccurate, there were those officers at the margins who exhibited substantial accuracy. For instance, the hit rate for weapons and contraband was 50% at the 95th percentile in 2014, compared to zero at the median, and reached 100% by

[189] *See* Andrew Caplin, Mark Dean & Daniel Martin, *Search and Satisficing*, 101 AM. ECON. REV. 2899, 2899 (2011); Winter, *supra* note 167, at 984–85.

[190] For instance, in 2006, on average there were 77 stops for every 100 Black males between 15 and 19 in New York City. Meares, *supra* note 59, at 339 (quoting Amanda Geller & Jeffrey Fagan, *Pot as Pretext: Marijuana, Race, and the New Disorder in New York City Street Policing*, 7 J. EMPIRICAL LEGAL STUD. 591, 624 (2010)).

[191] Meares, *supra* note 26, at 162.

2016, with the median hit rate remaining zero. The results do not explain whether these success stories were one-offs where officers simply got lucky, or if successful officers in one month maintain success in the next.

Looking over an officer's career, both analytic strategies show varying indicators of updating. A higher hit rate—more successful stops—in a prior month was positively correlated with a high hit rate—continued successful stops—in the current month, suggesting successful officers trend toward further success. But, there was no evidence that more stops in the current or prior months led to greater success—it does not appear officers are extrapolating from their past stop experiences more generally—and the length of time an officer appeared in the dataset had little practical significance on their success rates—indicating that time on the job may not mean much for whether an officer can accurately predict who to stop.

The empirical literature on street stops provides similarly unclear answers about how to improve accuracy. Ferrandino found that precinct-level success rates were unstable over time. Scoring on "efficiency" of stops, he found thirty-five precincts had lower scores, thirty-one had higher scores, and ten did not change between 2004 and 2010.[192] On the other hand, Minzner, following the careers of individual officers, found that success rates for probable cause searches were consistent over time, where successful and unsuccessful officers continued on those same tracks.[193]

Two different analytic strategies suggest that officers can and do learn month by month, at least on some metrics. Once officers reach hit rate *x*, those who improve are significantly more likely to reach *x+1* the following month. However, as the hit rate increases in the prior month, they reach a plateau from which it is very difficult to improve further. That asymptote may be elastic with respect to institutional pressures to increase police contacts that urge more frequent stops regardless of learning and hit rates. Given the generally low median hit rates, that peak may be fairly low, and subject to an external incentive regime that is untied to either learning or to the constitutional demands of reasonable suspicion.

How do officers learn? A possible answer may be that officers can glean from past search and frisk conduct information about what types of activity are indicative of an individual hiding weapons or contraband. This information may spread through networks of police officers, including both

[192] *See supra* note 107 and accompanying text.

[193] *See supra* note 106 and accompanying text.

positive learning but also norms of misconduct.[194] As officers conduct more searches and frisks, they gain more exposure to the possible circumstances that lead or do not lead to finding weapons and contraband. The same may be less so for arrests—regardless of what an officer finds on an individual, the officer may have reason to arrest that person, perhaps for a conduct-related offense like obstruction. However, officers do not appear to be learning simply by increasing the stop activity, or from increasing what they believe to be stops based on probable cause. Finally, as officers are proportionally exposed to and stop more Black and Latinx suspects, these officers' hit rates do not increase in accuracy. The programmatic nature of stop and frisk does not encourage officers to focus on individualized criminal conduct,[195] and this type of program, which prioritizes maximizing police presence and stops, does not coincide with updating.

During the study period, the institutional design of the NYPD—as well as other police departments across the United States[196]—prioritized the aggressive use of stop and frisk and quotas to proactively deter and prevent criminal activity and advance public safety, rather than just respond to it.[197] This model, however, can create perverse incentives[198] and result in poor oversight.[199] Here, the results indicate that this aggressive model did not

[194] Akshay Jain, Rajiv Sinclair & Andrew V. Papachristos, *Identifying Misconduct-Committing Officer Crews in the Chicago Police Department*, 17 PLOS ONE 1, 17 (2022); Cohen R. Simpson & David S. Kirk, *Is Police Misconduct Contagious? Non-trivial Null Findings from Dallas, Texas*, J. QUANT. CRIMINOLOGY (Jan. 12, 2022), https://doi.org/10.1007/s10940-021-09532-7 [https://perma.cc/U4QT-G59N]; George Wood, Daria Roithmayr & Andrew V. Papachristos, *The Network Structure of Police Misconduct*, 5 SOCIUS 1, 3–5 (2019); Marie Ouellet, Sadaf Hashimi, Jason Gravel & Andrew V. Papachristos, *Network Exposure and Excessive Use of Force: Investigating the Social Transmission of Police Misconduct*, 18 CRIMINOLOGY & PUB. POL'Y 675, 678–81 (2019).

[195] *See supra* notes 114–117 and accompanying text.

[196] Grunwald & Fagan, *supra* note 69, at 357–61; Fagan & Geller, *supra* note 85, at 53 n.7 (2015); *see, e.g.*, United States v. Weaver, 975 F.3d 94, 111 (2d Cir. 2020) (Livingston, J., dissenting) (noting that Judge Livingston would have found reasonable suspicion based on high crime area and defendant's conduct).

[197] *See* Harmon & Manns, *supra* note 83, at 53–57.

[198] Bronstein, *supra* note 129, at 553–56; Shaun Ossei-Owusu, *Police Quotas*, 96 N.Y.U. L. REV. 529, 535 (2021); Monica C. Bell, *Next-Generation Policing Research: Three Propositions*, 35 J. ECON. PERSPS. 29, 34–35 (2021); Richard H. McAdams, Dhammika Dharmapala & Nuno Garoupa, *The Law of Police*, 51 U. CHI. L. REV. 135, 147–49 (2015); Allison P. Harris, Elliott Ash & Jeffrey Fagan, *Fiscal Pressures and Discriminatory Policing: Evidence from Traffic Stops in Missouri*, 5 J. RACE ETHNICITY & POL. 450, 454–55 (2020).

[199] Stephen Clarke, *Arrested Oversight: A Comparative Analysis and Case Study of How Civilian Oversight of the Police Should Function and How it Fails*, 43 COLUM. J.L. & SOC. PROBS. 1, 4–10 (2009); Maggie Hadley, Note, *Behind the Blue Wall of Silence: Racial*

create an environment in which officers could effectively update and learn from prior stops. Rather than responding as rational actors to demands to efficiently target those likely to be involved in crime by learning and updating from prior stop activity, the results do not show consistent updating based on various stop factors.

It appears that, in a policing regime that demands increased policing activity as an indicator of productivity, without a concomitant demand for accuracy and updating, rationality goes out the window. Despite prioritizing the "quantity of enforcement activity,"[200] increased stop activity in the current and prior months was associated with *lower* hit rates. A policing regime that works on the principles of activity without considering returns threatens to neutralize law and rationality as a guidepost for behavior. While the regime imposes no costs for failures on the officers, these costs are externalized through a range of social and psychological harms on those subjected to "bad" or unproductive stops as a result of low hit rates.[201]

Following the *Floyd* litigation, NYPD eliminated quotas and hit rates improved as the number of stops declined, providing some evidence that a regime change can change behavior.[202] However, this is tempered by our results indicating that stops after *Floyd* class certification were negative predictors of hit rates for arrests and summons, and hit rates overall were only weakly significantly higher for those officers whose careers began after class certification. And there is evidence that these improvements in hit rates coincided with increasingly poor reporting of stops by officers,[203] suggesting a more complex institutional setting where incentives remain unaligned to rational stop behavior.

Courts, too, have reinforced this regime by placing substantial value on officers' own accounts of their conduct and experience as proxies for reasonable suspicion, rather than focusing on whether officers routinely

Disparities in NYPD Discipline, 53 COLUM. HUM. RTS. L. REV., 663, 668–82 (2022); Mary D. Fan, *Body Cameras, Big Data, and Police Accountability*, 43 L. & SOC. INQUIRY 1236, 1238–40 (2018).

[200] Bronstein, *supra* note 129, at 564.

[201] Jordan E. DeVylder, Hyun-Jin Jun, Lisa Fedina, Daniel Coleman, Deidre Anglin, Courtney Cogburn, Bruce Link & Richard P. Barth, *Association of Exposure to Police Violence with Prevalence of Mental Health Symptoms Among Urban Residents in the United States*, JAMA NETWORK OPEN 8–10 (Nov. 21, 2018), https://jamanetwork-com.turing.library.northwestern.edu/journals/jamanetworkopen/fullarticle/2715611 [https://perma.cc/7F29-GYGC].

[202] PETER L. ZIMROTH, SEVENTH REPORT OF THE INDEPENDENT MONITOR 14–21, 27–29 (2017) [ZIMROTH, SEVENTH REPORT], https://ccrjustice.org/sites/default/files/attach/2017/12/Monitor's.pdf [https://perma.cc/NV9D-HNJM].

[203] *See* Devereaux, *supra* note 159. Rates of reporting per stop are estimated now at less than 40%. ZIMROTH, NINTH REPORT, *supra* note 153, at 40–41.

effect productive stops and how they evaluate their prior stops, successes, and failures. Courts routinely reject suppression motions on the basis of officer experience and fail to clearly identify the parameters of lawful and unlawful searches and seizures.[204] Adding all this up suggests that rationality is neutralized while institutional demands become the forces that shape police activity.[205]

This has two important consequences. First, a constitutional consequence: it points to a disconnect between Fourth Amendment jurisprudence and police practice. Not only does the level of judicial analysis fail to align with the scale of stop and frisk in practice,[206] but the courts' tendency to defer to an officer's experience when evaluating the constitutionality and reasonableness of a stop does not reflect the reality that learning is not wholly connected to experience. We are not suggesting that *Terry* was wrongly decided or that the reasonable suspicion standard should be thrown out, regardless of its difficult administrability. Rather, our results should temper courts' willingness to defer to an officer's length of time on the job or experience simply patrolling and stopping individuals, without also considering the institutional pressures on officers to increase activity without increasing accuracy when evaluating the constitutionality of a stop.[207]

Officers, even experienced ones, have low hit rates and do not appear to learn from past stops based on the most exacting constitutional criteria. Our results indicate that, particularly for stops of non-white individuals, greater exposure of officers to those populations does not lead to greater accuracy, which is especially concerning given the NYPD's long history of discriminatory stop practices.[208] Institutional pressure to increase stops without pressure to increase positive returns from these stops will carry forward and multiply the racial biases and disparities in suspicion and stops

[204] Lvovsky, *supra* note 16, at 486–91; Siyl Liu & Esther Nir, *Mission Impossible? Challenging Police Credibility in Suppression Motions*, 33 CRIM. JUST. POL'Y REV. 584, 586–89 (2022); Esther Nir, *Empowering the Exclusionary Rule: Using Suppression Motion Data to Improve Police Searches and Searches in the United States*, 22 INT'L J. POLICE SCI. & MGMT. 96, 97–98 (2020). Scott W. Howe, *A Sixth Amendment Inclusionary Rule for Fourth Amendment Violations*, 54 CONN. L. REV. 613, 640–52 (2022) (arguing for replacement of the disfavored exclusionary rule with an "inclusionary rule" based on the right to trial under the Sixth Amendment).

[205] *See, e.g.*, Alex Chohlas-Wood, Marissa Gerchick, Sharad Goel, Aziz Z. Huq, Amy Shoemaker, Ravi Shroff & Keniel Yao, *Identifying and Measuring Excessive and Discriminatory Policing*, 89 U. CHI. L. REV. 441, 452–60 (2022).

[206] *See supra* notes 114–117 and accompanying text.

[207] Lvovsky, *supra* note 16, at 482, 491; *see also* Harris et al., *supra* note 198, at 4 (showing how the system of fines and fees operates to reward officers who cite suspects for purposes of municipal revenue generation rather than crime detection or control).

[208] *See supra* sections I.B.3.–4.

activity,[209] implicating not only the reasonable suspicion standard but disparate treatment and equal protection concerns as well. In sum, more experience may not mean fewer hunches as the *Terry* Court supposed, and if officers are not rational actors updating from prior stop activity, then courts can no longer consider reasonable suspicion to be "commonsensical."[210] Rather, in the context of institutional pressures to increase stop activity in a vacuum, what an officer believes to be reasonable may be irrational.

Second, a practical consequence: although police training traditionally has not focused on accuracy or hit rates,[211] recent work by Goel and colleagues have demonstrated how big data can help guide police officers' use of discretion.[212] There has been a significant shift in the NYPD's training and evaluation procedures post-*Floyd*, applying methods imposed by the court. However, the new procedures still do not focus on stop accuracy, only on the bases of the stops and officer conduct during stops.[213] The NYPD Police Commissioner has stated, "[l]arge numbers of arrests, summonses, and stops are not our goal."[214] If the outputs and outcomes or yield of stops are not the goal, increasing compliance with constitutional mandates by ensuring reasonable suspicion is not the only answer. If hit rates increase, officers can accomplish their public safety goals in fewer stops, imposing fewer burdens on civilians. It has been well documented that excessive stops have negative consequences,[215] and that too many stops that yield no results are an inefficient use of police manpower. In line with the processes explained in Goel's work, precincts can and should review hit rate data for individual officers over time, provide training to officers who trend downwards, and identify characteristics of sustained success. Officer incentives should be aligned with successful stops, not simply the overall number of stops. Moreover, given the racial disparities in stops and their

[209] *See* Grunwald & Fagan, *supra* note 69, at 359–62; Fagan & Geller, *supra* note 85, at 62–63.

[210] United States v. Lender, 985 F.2d 151, 154 (4th Cir. 1993) (explaining courts should "credit[] the practical experience of officers who observe on a daily basis what transpires on the street").

[211] *See supra* notes 108–113.

[212] Goel et al., *supra* note 11, at 232 ("[B]ig 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."); Goel et al., *supra* note 76, at 365 ("[W]e demonstrate that by conducting only the 6% of stops that are statistically most likely to result in weapons seizure, one can both recover the majority of weapons and mitigate racial disparities in who is stopped.").

[213] ZIMROTH, SEVENTH REPORT, *supra* note 202, at 14–21, 27–29.

[214] *Id.* at 2.

[215] *See* Meares, *supra* note 59, at 345–48.

outcomes, greater training and awareness of prior outcomes by race may help officers identify biases and allow their supervisors to identify potentially problematic stoppers. Once officers are given tools to update and learn from their past behavior, we can place more trust in *Terry*'s reliance on experience and training, and the functionality of the reasonable suspicion standard to regulate constitutional compliance. The alternative is to recalibrate the formation of suspicion and, as seems to be a productive guidepost,[216] apply a standard that substantially moves proactive stops toward a probable cause metric.

[216] Fagan, *supra* note 25, at 88 ("[S]hifting stops toward probable cause or behavioral indicia will shrink the stop circumstances that might otherwise be legally contested, reducing the burdens on trial and appellate courts. A shift in emphasis also creates a vocabulary and logic for internal audit, supervision, and regulation."); MacDonald et al., *supra* note 69, at 11 ("[T]his program may have been more productive if it placed more emphasis on probable cause stops more directly related to observable criminal activity.").