BIG DATA ANALYTICS, RISING CRIME, AND FOURTH AMENDMENT PROTECTIONS

Timothy J. Kraft[*]

TABLE OF CONTENTS

I. INTRODUCTION

The Fourth Amendment provides Americans with privacy protections,[1] yet these protections do not reach many personal communications on the public Internet. The potential implications for personal communications on the Internet—such as a blog post, a comment on a personal page, or a "tweet"—are legion. Data analytics tools combine demographic characteristics with known crime patterns in areas to reveal patterns and correlations. Combining data analytics with real-time social media constitutes a watershed moment for law enforcement practice and crime prevention. However, the lurking specter of law enforcement agencies trampling on the spirit of the Fourth Amendment

* J.D. & M.B.A. Candidate, University of Illinois at Urbana-Champaign, Class of 2017. I owe a debt of gratitude to God for His guiding hand, to my Mom, Adrienne, for her very patient and constant love, and to my Grandpa, Joe, who showed me how to be a man of integrity. In addition, I am grateful for the insight, effort, and patience of the Journal's editorial board in its undertaking of sharpening my prose.

1. *See* Terry v. Ohio, 392 U.S. 1, 8–10 (1968) (providing that *Terry* stops have the legal sanction of the highest court provided they are executed within the bounds of Fourth Amendment protections).

is equally palpable. Currently, for the purposes of conducting a *Terry* stop, law enforcement may be able to justify reasonable suspicion by combining superficial observations of a person with personal Internet posts or data analytic outputs alone. However, justifying *Terry* stops on such tenuous inferences curtails some of the protections put in place by the Founding Fathers. Consequently, there needs to be reinforcement, preferably from Congress, of the Fourth Amendment's reasonable suspicion test for *Terry* stops where law enforcement makes statistical inferences based on data analytics and social media posts.

Although the U.S. Supreme Court has amassed an extensive body of law regarding the Fourth Amendment,[2] it has not yet explicitly laid down a rule regarding law enforcement's use of data analytics to support a *Terry* stop when an officer combines observations of a person's behaviors with descriptive statistics and correlations.[3]

The Fourth Amendment inserted procedural safeguards into criminal investigations.[4] Generally, a search or seizure effected by law enforcement must be reasonable as demonstrated by the existence of probable cause and a judicial warrant.[5] However, several tailored exceptions to the Fourth Amendment's requirement of probable cause alleviate the requirement for law enforcement to obtain a warrant.[6] There are two Fourth Amendment concerns implicated by the use of data analytics and real-time social media content. The first concern, and a focus of this Note, is law enforcement's accumulation and use of data as it relates to making statistical inferences for *Terry* stops in cases of marginal reasonable suspicion. The second concern is the search or seizure of publicly viewable data on social media that, by virtue of it being knowingly exposed to the public, can be gathered by law enforcement without any suspicion at all. This is distinguishable from "semi-protected" data that the user has taken some steps to keep private but that others may still access.

Limiting the scope of the sources that comprise reasonable suspicion presents problems for law enforcement in the era of big data, facial recognition software, and rapid analytics. For example, suppose police agencies stored preselected data from criminal arrests in a shared database.[7] These data points

2. *See* Shima Baradaran, *Rebalancing the Fourth Amendment*, 102 GEO. L.J. 1, 11–12 (2013) (detailing the subject matter of the extensive U.S. Supreme Court decisions on the Fourth Amendment).

3. *See* Kelly K. Koss, *Leveraging Predictive Policing Algorithms to Restore Fourth Amendment Protections in High-Crime Areas in a Post-Wardlow World*, 90 CHI.-KENT L. REV. 301, 301 (2015) ("The United States Supreme Court has yet to hear a case addressing issues related to law enforcement's use of data-driven tips from predictive policing software.").

4. *See generally* Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 329 (2015) (discussing how the Fourth Amendment is applied in the law enforcement context and the requirement of reasonable suspicion as a procedural safeguard).

5. *See* U.S. CONST. amend. IV ("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.").

6. *See generally Warrantless Searches and Seizures*, 40 GEO. L.J. ANN. REV. CRIM. PROC. 44, 44–45 (2011) (listing the exceptions that have been hewn from the requirement of reasonableness embodied in the general requirements of probable cause and a judicially issued warrant).

7. *See* Ferguson, *supra* note 4, at 330 (articulating the existence of law enforcement databases and

could include such details as location of arrest, time of arrest, race of arrestee, arrestee's clothing, suspicious activity precipitating detention, and crime with which arrestee was charged.[8] The local law enforcement agency may augment these databases with automatically updating profiles of convicts, parolees, and persons of interest.[9] At first blush, such law enforcement activity may appear dystopian, but gang units in urban areas already utilize the same or similar profiles.[10] Over time, these data points accumulate, giving law enforcement officers a robust base of knowledge easily distributed to each patrolling squad car.[11] Each police officer then may assemble reasonable suspicion using correlative statistics to make the inference that the observed individual meets the criteria of a person who is committing or soon will commit a crime when, in a reality divorced from big data analytics, such reasonable suspicion may be sorely wanting.[12]

Despite the Orwellian fears of law enforcement officers incorrectly profiling innocents as criminals, several policy considerations arguably provide a foundation for supporting state use of data analytics, real-time social media, web crawlers, and data gathering tools.[13] The potential for a false-positive identification or racial profiling is significant, but apologists for big data in police work argue that, as the data sets grow, identifications of potential criminals will become increasingly accurate.[14] Other defenders of integrating data analytics into law enforcement point to the efficient use of public tax revenues as each dollar spent can be targeted at high crime areas, at the time of day when crimes are most likely to be committed, and at suspects most likely to commit the crimes.[15] Hypothetically, law enforcement accountability would improve as well, as officers would have the opportunity to point to the specific descriptive statistic or correlation that augmented their observations giving rise to reasonable suspicion, which then justified them stopping a member of the public.[16]

---

other sources of information that can be used to manufacture reasonable suspicion).

8. *Id.*

9. *See, e.g.*, Danielle Gordon, *Police Database Profiles Activist as Gang Member*, CHI. REP. (Sept. 25, 2007), http://chicagoreporter.com/police-database-profiles-activist-gang-member/ (providing an example of updating police records).

10. *Id.*

11. *See, e.g.*, *Most Wanted*, L.A. POLICE DEP'T, http://www.lapdonline.org/most_wanted (last visited Mar. 11, 2017) (providing an example of a most wanted list distributed to each patrolling police officer); *see also* Michele Coppola, *Minnesota Police Department Finds Ways to Embrace Technology*, TECHBEAT, Jan.–Feb. 2014, at 7, 9, http://justnet.org/InteractiveTechBeat/eTECHBEAT/201401/pdf/eTechBeat_Winter_Issue_2014.pdf ("We want to ensure information is shared so that everyone has a good idea of what is going on. All the information on the [department] intranet is accessible from the squad car and department computers.").

12. *See* Koss, *supra* note 3, at 302–03 (describing how a law enforcement office could utilize statistical outputs to claim a reasonable suspicion to stop a citizen).

13. Ferguson, *supra* note 4, at 388–97.

14. *See generally* Rachael King, *IBM Analytics Help Memphis Cops Get "Smart"*, BLOOMBERG (Dec. 5, 2011, 9:30 PM), http://www.bloomberg.com/news/articles/2011-12-05/ibm-analytics-help-memphis-cops-get-smart (discussing the Memphis Police Department's use of big data technology to drive crime rates down).

15. *See generally* Ferguson, *supra* note 4, at 394 (articulating an efficiency defense for the use of data analytics vis-à-vis scarce department resources).

16. *Id.* at 393.

Part II of this Note details the development of jurisprudence regarding reasonable suspicion that is generated by data analytics and augmented by real-time Internet posts, which is subsequently used to justify *Terry* stops. Part III analyzes the judicial framework of current Fourth Amendment law and its coalescence with the evolution of big data analytics. It concludes by highlighting some of the challenges and unresolved questions posed by law enforcement's use of big data analytics in combating crime. Lastly, Part IV proposes a new balancing test that encompasses investigatory procedure, big data, and third-party records, and operates in concert with the American Bar Association's (ABA) view as set forth in its Criminal Justice Standards on Law Enforcement Access to Third Party Records (LEATPR Standards).[17] Specifically, Congress should pass legislation that requires a *Terry* stop supported by reasonable suspicion to be based on more than data analytics and correlations to be sufficiently particularized and individualized. The proposed balancing test should weigh the individual's objective expectation of privacy in their information—per the ABA's LEATPR Standards, together with certain plus factors,[18] such as proactive actions to keep transmitted Internet communications private[19]—against a reasonable suspicion standard for each successive Internet search focused on personally identifiable information that is supported by massive data set analytics.[20]

II. BACKGROUND

A. Data Gathering and Statistical Reasonableness

An individual's right to be secure in one's person and property was enshrined in the sacrosanct words of the Fourth Amendment: "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 [w]arrants shall issue, but upon probable cause . . . ."[21] Thus, the Constitution itself mandates that every search or seizure made by a state actor be reasonable.[22]

Several exceptions to the standard of probable cause coupled with a warrant have been carved out since 1789 to include "investigatory stops [*Terry* stops], . . . searches incident to a valid arrest, seizures of items in plain view, . . . consensual searches, . . . [and] seizures of abandoned property," among

17. *See generally* STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS (AM. BAR ASS'N, amended 2013) (describing the ABA's position on a reasonable expectation to privacy that attaches to certain Internet communications to third parties).

18. *See* William E. Kovacic et al., *Plus Factors and Agreement in Antitrust Law*, 110 MICH. L. REV. 393, 395–96 (2011) (articulating the use of plus factors as circumstantial evidence in antitrust law).

19. *See generally* Horton v. California, 496 U.S. 128, 133 (1990) (reasoning that individuals have an interest in privacy, but when evidence is left in plain view no proactive actions have been taken to keep the item private).

20. *See generally* Ferguson, *supra* note 4, at 329 (discussing how the Fourth Amendment is applied in the law enforcement context and the requirement of reasonable suspicion).

21. U.S. CONST. amend. IV.

22. *Id.*

others.[23]

As the Supreme Court pointed out in *Terry v. Ohio*, law enforcement must have reasonable suspicion to perform an investigatory stop of a suspect.[24] Under the law, reasonable suspicion is a product of the totality of the circumstances that is tailored to the individual in the context of place and time.[25] Professor Ferguson, an influential legal scholar in the field of technology and Fourth Amendment law, argues that the reasonable suspicion doctrine is a "small data" doctrine where small data is defined as discrete, readily observable facts about the specific individual.[26] As data processing has increased in efficacy and data storage has improved in cost and efficiency, law enforcement can access particularized information about an observed individual through the Internet, real-time social media posts, and databases.[27] Moreover, local and national agencies can glean important insights—such as whether the unaware citizen under law enforcement's observing eye is statistically likely to engage in a crime at the time and place of observation—through the descriptive statistics, heat maps, and correlations that result from data mining programs.[28]

Personal observations, once the sole content of reasonable suspicion analyses, can now be augmented by a broad swath of data on an individual via Internet posts, other web-based activities, and databases concerning the individual's characteristics, location, and criminal history.[29] The robust data available to the law enforcement officer may indicate that a particular individual, though not currently engaged in criminal activity, has a high probability of engaging in a criminal act. It follows that the officer may make a statistical inference based on the preceding data analyses that mirrors what the law deems an individualized suspicion that—when coupled with a few discrete articulable observations such as loitering—may rise to the level of reasonable suspicion.[30] In this situation, the protection supposedly afforded by the requirement of reasonable suspicion is substantially weakened. Although the law enforcement officer in the above example may be suspicious of a person's criminal record, recent social media activity, and her particular demographic correlation with analytic outputs, an officer's ability to search or seize a person without having personally observed reasonably suspicious behavior seriously encroaches on the protections afforded by the Fourth

23. *Warrantless Searches and Seizures*, *supra* note 6, at 44–45, 141.

24. *See* Terry v. Ohio, 392 U.S. 1, 21 (1968) ("[I]n justifying the particular intrusion the police officer must be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant that intrusion.").

25. *See* United States v. Arvizu, 534 U.S. 266, 273 (2002) (mandating a "totality of the circumstances" analysis to determine if there is a sufficient basis for suspecting criminality).

26. *See* Ferguson, *supra* note 4, at 329 (articulating his idea that reasonable suspicion is a small data doctrine based on "discrete facts, limited information, and little knowledge about the suspect").

27. *Id.* at 330.

28. Koss, *supra* note 3, at 302–03.

29. Ferguson, *supra* note 4, at 330.

30. *See id.* at 331 ("This new reality simultaneously undermines the protection that reasonable suspicion provides . . . and potentially transforms reasonable suspicion into a means of justifying those same stops.").

Amendment.[31] Yet, proponents of big data analysis argue that the reasonable suspicion standard is made much more robust, and more reasonable, by a data set describing behavior patterns beyond the readily observable actions of the individual.[32]

The legal problem arises when the indexing of third-party websites (i.e., blogs, personal social media profiles, etc.,) facilitates data gathering, thus permitting law enforcement to construct digital identities from the vestiges of one's activity on the web.[33] No court has established a rule concerning what test applies to content posted by private citizens to third-party websites in regard to whether their content is properly understood to be in plain sight or subject to a new judicial test when used for data analysis and statistical inferences. However, courts have long held that a person who provides information to a third party forfeits any expectation of privacy attaching to the transmitted information, thus explaining why this area of apparently settled law has not yet been revisited.[34]

In lieu of a cogent doctrine applicable to law enforcement's access to third-party Internet websites in pursuing criminal investigations or seeking to prevent crime, the ABA has set forth new standards—in its Criminal Justice Standards on Law Enforcement Access to Third Party Records—that encompass this species of investigatory data gathering.[35] As Professor Ferguson points out, "[t]he value in 'third party records' is information—masses of revealing information."[36] Law enforcement has an incentive to prevent crime by obtaining data sets to produce confidence intervals for the most probable personal characteristics of an offender as well as the likely time and place of a particular crime.[37] In contrast, individuals can reply by pointing to the Fourth Amendment and its seemingly stalwart prohibition against law enforcement probing into their papers,[38] if papers may be construed to include their Internet activity on a third-party website when some concerted action has been taken to maintain the privacy of that transmitted information. The ABA LEATPR Standards chart a middle path to balance the competing interests of law enforcement (whose duty it is to protect the general public) and private citizens (whose right it is to "be secure in their persons and property").[39] Further, the Standards propose to limit the type of records law enforcement can

31. *Id.*

32. Koss, *supra* note 3, at 302–03.

33. *See* Junghoo Cho, Crawling the Web: Discovery and Maintenance of Large-Scale Web Data (Nov. 2001) (unpublished Ph.D. dissertation, Stanford University), http://oak.cs.ucla.edu/~cho/papers/cho-thesis.pdf (discussing how web crawler software can retrieve data from third-party websites and periodically update their data).

34. *See* Smith v. Maryland, 442 U.S. 735, 743–44 (1979) (discussing the third-party doctrine and subsequent data collection by the U.S. Government from third-party telecom pen registry).

35. *See* Andrew Guthrie Ferguson, *Big Data Distortions: Exploring the Limits of the ABA LEATPR Standards*, 66 OKLA. L. REV. 831, 831–33 (2014) (providing an overview of the ABA LEATPR Standards and their applicability to law enforcement's use of big data and data mining).

36. *Id.* at 831.

37. *See* Koss, *supra* note 3, at 310 (examining how law enforcement can leverage statistical outputs over a risk terrain map to prevent crime).

38. U.S. CONST. amend. IV.

39. *Id.*; *see* Ferguson, *supra* note 35, at 861–65.

obtain, but they permit the use of such records in the aggregate and erect protections to insulate personally identifiable data.[40]

B. Database Building and Data Mining

Good police work has always required the kind of inferences drawn between crime, location, time, and perpetrator, and these inferences are now being suggested by large-scale database analytics.[41] Data mining and the market for data sets has become so prevalent in American life that industries have begun providing their consumer data to law enforcement whenever an agency simply asks.[42] The law enforcement agency may then augment the commercially available data set with its own records of criminal behavior.[43] The amalgam of commercial data and law enforcement data provides a firm foundation to tailor a data query to high-traffic websites, personal web pages, and other sites with potentially incriminating content that is a boon for investigators but inimical to the privacy of the individual.[44]

These massive data sets are being leveraged by national and local law enforcement agencies through cost-effective data storage, increased processing power, and precise search parameters, which theoretically results in less crime through greater efficiency and efficacy in police work.[45] Despite recent attention given to big data analytics by the press, law enforcement's use of computers and databases to perform data analysis and even predict criminal behavior has been a common practice for many years.[46] Before computers, law enforcement would use pushpins on a map to derive insights from data and build dossiers on suspected criminals while observing individual behaviors on their "beat."[47] Presently, new technology has expanded the boundaries of investigatory practice through web crawler software, database building, and data mining.[48]

A web crawler is simply a computer program that seeks out specified web pages using search terms and targeted algorithms in order to catalogue them and the data they contain.[49] The crawler also periodically updates the stored

40. *Id.*

41. *See* Chris Jay Hoofnagle, *Big Brother's Litter Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595, 595–96 (2004) (depicting how law enforcement is using data analytics).

42. *See* Stephen Rushin, *The Judicial Response to Mass Police Surveillance*, 2011 U. ILL. J.L. TECH. & POL'Y 281, 289 (2011) (describing the voluntary flow of consumer data to law enforcement).

43. Fred H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435, 457 (2008).

44. *Id.* at 444 ("[T]he [Investigative Data Warehouse] contains more than 659 million records, which come from 50 FBI and outside government agency sources . . . [and] agents and analysts who use the system average one million queries a month.").

45. Ferguson, *supra* note 35, at 835–36; *see also* Koss, *supra* note 3, at 312.

46. Cate, *supra* note 43, at 438.

47. *See* Koss, *supra* note 3, at 302 (illustrating the use of pushpins as one method to derive new insights from data).

48. *See generally* Ferguson, *supra* note 4, at 329–30 (describing how new technology has led to reforms in investigatory practice in law enforcement agencies).

49. *See* Cho, *supra* note 33, at v (defining web crawler software).

web page data.[50] The periodic updates enable nearly real-time analysis of changes in the underlying data.[51] Web crawlers can catalogue any web page including external, third-party websites, such as Facebook, Twitter, Instagram, the *Huffington Post*, and others, limited only by the knowledge one has of websites.[52] Because computing power is limited, law enforcement agencies need to prioritize the data they want to collect as well as the key websites from where they would like to collect it.[53]

Web crawlers collecting data on third-party websites in addition to the periodic updates of those website's content (i.e., new posts, new blogs, new pictures, etc.) provide law enforcement with the massive data sets necessary to engage in data mining.[54] Moreover, "the quantity of the world's recorded data has doubled every year. At the same time, the computing power necessary to store, access, and analyze these data has increased geometrically, at increasingly cheaper cost."[55] Data mining has several manifestations, but its purpose is always the same—to glean new and possibly hidden insights from data.[56] Consider target-driven data mining and event-driven data mining. Target-driven data mining uses algorithms and search terms to focus outputs on particular individuals—either to determine past behavior or to determine the statistical likelihood for future criminal acts.[57] Event-driven data mining utilizes the massive data set accumulated to create patterns of behavior or activity that law enforcement may use to predict future patterns across locations, demographics, and time.[58] Thus, the federal and local governments have incrementally expanded the scope of data collection operations that allow for tracking of public records (such as prior arrests) and other websites where personal data may be collected for law enforcement purposes.[59]

Given the advances in technology, it is achievable for law enforcement agencies to collect new data whenever a private citizen posts a photo to Instagram, types out a 140-character tweet, comments on a blog, or publishes a status update on any third-party platform.[60] Courts have yet to conclusively determine whether such personalized data is properly construed as protected under the Fourth Amendment, subject to government data mining given its public exposure, or lies somewhere in between the two poles; such as in cases when additional steps are taken to protect the data.[61] What is clear is that data

50. *Id.*
51. *Id.* at 57.
52. *Id.* at 7.
53. *Id.*
54. *See generally* Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317 (2008) (describing the U.S. Government's efforts in using massive data sets to conduct data mining).
55. *Id.* at 317.
56. Ferguson, *supra* note 35, at 836.
57. Slobogin, *supra* note 54, at 322.
58. *See id.*; *see also* Koss, *supra* note 3, at 309–10.
59. *See* Slobogin, *supra* note 54, at 323 (describing other government programs analyzed by the U.S. Government Accountability Office).
60. *See generally* Ferguson, *supra* note 35, at 835, 837–38 (describing the ability of government to sift through massive amounts of data that includes data transmissions to public and private organizations).
61. *See* Koss, *supra* note 3, at 301–02.

mining of this sort can not only map the risk of crime to a location on a map and a time of day, but can also alert law enforcement to anomalous behavior in individuals without a criminal record.[62] Shifts in the way the world deals with data has spurred on this tectonic change in the world of law enforcement.[63] The information collectable from social networks and databases makes such data mining activities fruitful for law enforcement.[64]

III. ANALYSIS

Law enforcement agencies across local, state, and federal jurisdictions are augmenting the traditional "beat" approach to police work with big data mining and analytics to reduce crime and provide more effective patrols.[65] While an increased law enforcement presence at the locations most susceptible to a particular crime at the time when criminal activity is most often perpetrated may be sufficient to reduce crime, police must nevertheless adhere to the Fourth Amendment's circumscriptions on investigatory procedure.[66] Yet, the standard for making a lawful investigatory stop, as articulated by the Supreme Court, remains the reasonable suspicion standard.[67] The question analyzed here is multifaceted. First, is personal data in the form of posts, pictures, comments, or blogs on the Internet subject to inclusion in law enforcement agencies' databases when certain actions are taken to maintain privacy? If so, can these data points give rise to the reasonable suspicion necessary to make an investigatory stop absent some additional contemporaneous observations? If individual datum may not, might the same datum coupled with other bits of data give rise to the necessary reasonable suspicion? If so, is there a point at which the aggregation of the data implicates the Fourth Amendment? Implications for privacy protections in property, in papers, and in persons arising from the Fourth Amendment and as developed by the Supreme Court are widespread.[68]

62. Ferguson, *supra* note 35, at 833 ("[L]aw enforcement will be able to use predictive analytics to discover unusual patterns that might signal criminal activity from currently unknown individuals.").

63. *See* Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1186 (2002) (arguing it is the aggregation of the individual datum into an interconnected mass that makes information more powerful).

64. *Id.* at 1185–86.

65. *See* Mike Wheatley, *FBI Uses Big Data & Crowdsourcing to Hunt the Boston Bomber*, SILICONANGLE (Apr. 17, 2013, 12:10 AM), http://siliconangle.com/blog/2013/04/17/fbi-uses-big-data-crowdsourcing-to-hunt-the-boston-bomber/ (describing the FBI's use of big data in investigating the 2013 Boston marathon bombing); *see also* Gordon, *supra* note 9 (providing an example of database data use in a local context).

66. *See* Katz v. United States, 389 U.S. 347, 358–59 (1967) (holding that the Fourth Amendment requires a warrant for searches and seizures).

67. *See* Terry v. Ohio, 392 U.S. 1 (1968) (holding that the Fourth Amendment applies to a citizen on the streets as well as at home or elsewhere).

68. *See* Ferguson, *supra* note 4, at 373–76 (describing in broad strokes some of the implications of big data on reasonable suspicion).

A. The Big Data World

Law enforcement revolves around the tripartite core of "observation, investigation, and prediction."[69] Integrating big data analytics does not alter the primary methods of approaching police work through observations of citizens in neighborhoods where law enforcement agents work and live and through the investigations of crimes. Rather, traditional methods are augmented by making statistical inferences based on data analytics to prevent future crimes.[70] Data mining, analytics, and web crawler programs provide new tools to enhance the efficacy and efficiency of traditional police work.[71]

Consider the following hypothetical derived from the facts of *Ornelas v. United States*.[72] A law enforcement officer observes a young man lingering outside a convenience store at 9:24 p.m. The officer knows there has been a string of overdoses in the neighborhood resulting from a particular kind of drug being sold, but no arrests have been made. Performing a search from her patrol vehicle of the law enforcement database to investigate the young man—using individual characteristics such as approximate weight, height, race, and identifying scars and tattoos—the officer obtains a name, an arrest record including petty theft, and a last known address from what she believes is a positive identity match of the young man.[73] The officer continues her investigation by accessing the most recent social media activity of the young man that, based on her experience, *suggests* to her he is associated with a known gang in the area.[74] Additionally, the officer gleans insights from the data amassed on arrests from drug-related incidents in this neighborhood over the last twenty years using a law enforcement database.[75] Together, these data points, coupled with analytical software, suggests there is a 67% chance a person selling drugs matches the description of the young man. Professor Leipold highlights the complication of integrating data analytics at this juncture: namely, if these data points suggest to the patrolling officer that the observed individual is selling drugs right now, then there is no *Terry* violation.[76] Conversely, if the data merely suggests that a person matching these characteristics is selling drugs, then a Fourth Amendment problem arises because the "individualized suspicion" results from statistical correlations rather than contemporaneous observations.[77]

In isolation, any one of these factors may be insufficient to give rise to the reasonable suspicion necessary to execute an investigatory stop. Taken

69. *Id.* at 377.

70. *Id.*

71. *See* Koss, *supra* note 3, at 306 (describing the results in New York City from the use of big data and statistical outputs in decreasing crime).

72. Ornelas v. United States, 517 U.S. 690, 691–93 (1996).

73. Professor Ferguson suggests such data points are readily obtainable in his example involving a robbery suspect. Ferguson, *supra* note 4, at 330.

74. *Id.*

75. *Id.*

76. Interview with Andrew D. Leipold, Edwin M. Adams Prof. of Law, Univ. of Ill. Coll. of Law (Jan. 19, 2017).

77. *Id.*

together in a totality-of-the-circumstances inquiry, the officer arguably has obtained reasonable suspicion to conduct an investigatory stop, as the Court reasoned in *Ornelas*.[78] While there is the possibility that the investigatory stop may yield no contraband (i.e., a false positive), the data available to the patrol officer makes her reasonable suspicion more likely to result in an effective stop.

The above example illustrates how big data utilization, coupled with web crawlers and predictive analytics, simultaneously lowers the bar for reasonable suspicion but deepens the reasonableness of that suspicion.[79]

Big data is the use of massive data sets generated from the collection of disparate sources that, when aggregated, can be mined and used to conduct deeper levels of analysis.[80] This deeper analysis can reveal patterns, correlations, and statistically significant predicates of observable phenomena.[81] In order to successfully leverage big data, an organization must consider the three Vs—volume of data, variety of data sources, and velocity of data refreshing.[82] Volumes of data accumulate from individual interactions with the digital environment that leave behind digital records.[83] A person's digital activities create distinct records that, when aggregated by a data collecting organization (e.g., telecoms, social media websites, search engines, etc.), become a digital identity that can imbue law enforcement agencies with a powerful new set of pushpins for criminal investigations.[84]

Americans' digital identities are protected by both federal and local laws that require judicial process to view, but as Professor Ferguson points out, the "same information [can be obtained] through data aggregating services" that circumvent statutory protections.[85] Coupled with the interagency National Crime Information Center (NCIC) database, law enforcement agencies can glean investigatory advantages by accessing both aggregated data and, via the NCIC, personal addresses and demographic information.[86] Together, these data points provide personal characteristic information as well as personal criminal histories from which law enforcement can focus the scope of pending investigations to particular persons of interest.[87] Returning to the hypothetical

78. Ornelas v. United States, 517 U.S. 690, 700 (1996).

79. Ferguson, *supra* note 4, at 349–50.

80. *See* JAMES MANYIKA ET AL., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY (FULL REPORT) 1–3 (June 2011), http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation (discussing the term "big data," its subjective nature, and its size relative to discrete, small datum).

81. *Id.*

82. *See* JULES J. BERMAN, PRINCIPLES OF BIG DATA: PREPARING, SHARING, AND ANALYZING COMPLEX INFORMATION, at xv (2013) (articulating the "three Vs of big data" concept).

83. *See* Nicholas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 397–98 (2012) (examining Justice Sotomayor's "breadcrumbs" approach to big data and how small discrete datum can be aggregated into "big data").

84. *See generally* Lior Jacob Strahilevitz, *Reputation Nation: Law in an Era of Ubiquitous Personal Information*, 102 NW. U. L. REV. 1667, 1670–74 (2008) (describing how readily ascertainable personal information enables individuals and organizations to grasp, in a broad way, another person's identity).

85. Ferguson, *supra* note 4, at 360.

86. *National Crime Information Center (NCIC)*, FBI, https://www.fbi.gov/services/cjis/ncic (last visited Mar. 11, 2017).

87. *Id.*; Ferguson, *supra* note 4, at 360.

above, law enforcement can use this amalgam of information to justify a "*Terry* stop" of a suspect whose physical characteristics and criminal history may justifiably arouse a reasonable suspicion.[88] These uses of diverse streams of data illuminate the "second V," namely, variety of data sources.[89]

When diverse information streams are unified in a database, the power of the pooled information is multiplied.[90] For example, the Department of Homeland Security, in an effort to stem the possibilities of terrorist attacks, built the Multi-State Anti-Terrorism Information Exchange Program (MATRIX).[91] MATRIX enabled local law enforcement to access various information on private citizens such as criminal history, financial information, vehicle information, driver's license information, concealed weapons permits, professional licenses, voter registration records, and other information.[92] In an internal audit of the MATRIX program, the Department of Homeland Security revealed in a public release that "only 2.6% of the cases investigated over the course of the MATRIX pilot project were related to terrorism."[93] The other 97.4% of cases that incorporated the MATRIX program were criminal investigations related to narcotics, homicide, and fraud, among other non-terrorism-related crimes.[94] Law enforcement's use of the MATRIX pilot program affirms big data's value to investigations when aggregated on one platform.[95] However, the Department of Homeland Security voluntarily disbanded the program amid privacy concerns from activist groups such as the American Civil Liberties Union and not because of complications arising from Fourth Amendment law.[96]

Despite its early demise, MATRIX's legacy continues to affect modern law enforcement via the public-private partnerships that followed in its wake.[97] The Federal Bureau of Investigation (FBI) devotes significant resources to supply its agents and analysts with current data collected from several public and private organizations.[98] These public-private partnerships yield bountiful data harvests that are collected and maintained by the FBI's Criminal Justice Information Services Division (CJISD).[99] Private sector companies have expanded their data product line to provide purportedly private information on

88. *See generally* Terry v. Ohio, 392 U.S. 1 (1968) (providing the legal basis of reasonable suspicion for *Terry* stops).

89. BERMAN, *supra* note 82.

90. *See generally* Ferguson, *supra* note 4, at 360 (discussing the improvements to investigations when information is networked and aggregated in databases).

91. U.S. DEP'T OF HOMELAND SEC., REPORT TO THE PUBLIC CONCERNING THE MULTISTATE ANTI-TERRORISM INFORMATION EXCHANGE (MATRIX) PILOT PROJECT (Dec. 2006), http://www.dhs.gov/xlibrary/assets/privacy/privacy-matrix-122006.pdf (last visited Mar. 11, 2017).

92. *Id.*; Ferguson, *supra* note 4, at 361.

93. U.S. DEP'T OF HOMELAND SEC., *supra* note 91, at 2.

94. *Id.*

95. *Id.*

96. *Id.* at 4; Anita Ramasastry, *Why We Should Fear the Matrix: The "Multistate Anti-Terrorism Information Exchange" Program Threatens to Revive Total Information Awareness*, FINDLAW (Nov. 5, 2003), http://writ.news.findlaw.com/ramasastry/20031105.html.

97. Ferguson, *supra* note 4, at 362–63.

98. *See* Cate, *supra* note 43, at 442–44 (describing the FBI's public-private partnership's data mining cooperation).

99. *Id.*

customers and end users to government agencies.[100] Data gathered by private companies, like Google, are often more robust, more probative, and more revealing than the information typically held in the public domain.[101] Thus, private sector data collection has the potential to duplicate and replace what now-defunct public sector programs, such as MATRIX, provided to public law enforcement agencies.[102]

B. Supporting Reasonable Suspicion with Social Media

Law enforcement agencies can augment reasonable suspicion by harvesting data in the public domain.[103] As a regulatory matter, there exist relatively few prohibitions from law enforcement "accessing and collecting personal data" that accumulates from an individual's activities in the public domain of the Internet.[104] In *Katz v. United States*, the Supreme Court held that law enforcement listening to the conversation of a private citizen in a public telephone booth was a search and seizure under the Fourth Amendment because his expectation of privacy within the phone booth was protected.[105] Justice Harlan, in his concurring opinion, articulated that the expectation of privacy must be a reasonable one in order to be constitutionally protected.[106] However, the reasonable person standard as applied to one's expectation of privacy does not map so neatly onto individuals who have already come into contact with law enforcement and the court system because of the already existing record they have.[107] Moreover, the Supreme Court strictly limited the reach of Fourth Amendment protections by articulating the rule that a person does not have a reasonable expectation of privacy in anything he or she knowingly divulges to the public.[108] Thus, any information in the public view could be used in court as evidence against a defendant who had unwittingly revealed personal details when that person did not take any proactive actions to preserve the privacy of the information.[109]

Coolidge v. New Hampshire elaborated on the foundation of *Katz* and specified that the Fourth Amendment does not protect information a person

100. *See* Terry, *supra* note 83, at 391 (examining private companies' sale of big data as facilitating government surveillance); *see also* Andy Greenberg, *U.S. Government Requests for Google Users' Private Data Jump 37% in One Year*, FORBES (June 17, 2012, 11:01 PM), http://www.forbes.com/sites/andygreenberg/2012/06/17/u-s-government-requests-for-google-users-private-data-spike-37-in-one-year/#483743622c02 (providing an example of the U.S. Government's ability to obtain Google's end-user data).

101. *See* Terry, *supra* note 83, at 389–91 (discussing one company, Acxiom, which has roughly 1,500 data points on each person of interest).

102. *See* Candice L. Kline, Comment, *Security Theater and Database-Driven Information Markets: A Case for an Omnibus U.S. Data Privacy Statute*, 39 U. TOL. L. REV. 443, 447–449 (2008) (elaborating on the interconnected nature of data aggregation across corporate industries that can be packaged and sold to government agencies); Ferguson, *supra* note 4, at 362–64.

103. Ferguson, *supra* note 4, at 373.

104. *Id.*

105. Katz v. United States, 389 U.S. 347, 356–58 (1967).

106. *Id.* at 360–61.

107. Ferguson, *supra* note 4, at 373–75.

108. *Katz*, 389 U.S. at 351.

109. *Id.*

reveals to a third party that the third party then reveals to law enforcement.[110] The rationale is one assumes the risk that the information divulged to the third party may be revealed to others and thus does not have a reasonable expectation to privacy.[111] The third party to whom an individual reveals potentially incriminating information can also be a corporation and includes Internet search providers.[112] While a growing number of academics advocate for the Court to revisit this doctrine, current Fourth Amendment law does not extend to protect these voluntary disclosures of information.[113]

Some data is protected and has been categorically restricted from arbitrary government access.[114] Data disclosed to third parties that is protected includes health data, financial data, and the content of e-mails and telephonic communications.[115] Legislation like the Electronic Communications Privacy Act of 1986 and the Stored Communications Act endeavor to prevent the unauthorized access and use of electronic communications, but they do not extend to electronic communications in the public domain.[116] Voluntary consumer protection policies drafted by companies like Facebook and Google likewise provide some measure of protection of information posted, blogged, queried, and written on third-party websites.[117] The extent of such voluntary protections is questionable when weighed against the value customer data offers to companies that create revenue streams by offering personal data as a product to other firms.[118]

Individuals who purposefully publish personal information on the Internet without privacy safeguards reveal information to third parties that is not protected by the Fourth Amendment.[119] This information can be coupled with new technologies to give rise to a sufficient number of factual antecedents to justify a reasonable suspicion.[120] New technology called the Domain Awareness System (DAS) implemented by the New York Police Department (NYPD) and developed by Microsoft enables law enforcement officers to "blend[] and analyz[e] realtime data gathered from roughly 3,000 civic closed-circuit cameras . . . and license plate readers."[121] DAS technology together

110. Coolidge v. New Hampshire, 403 U.S. 443, 487–490 (1971).

111. Richard A. Epstein, *Privacy and the Third Hand: Lessons from the Common Law of Reasonable Expectations*, 24 BERKELEY TECH. L.J. 1199, 1222 (2009).

112. Ferguson, *supra* note 4, at 373–75.

113. *Id.*

114. Erin Murphy, *The Politics of Privacy in the Criminal Justice System: Information Disclosure, the Fourth Amendment, and Statutory Law Enforcement Exemptions*, 111 MICH. L. REV. 485, 487 (2013).

115. Ferguson, *supra* note 4, at 375–76.

116. 18 U.S.C. §§ 2701–12 (2012).

117. *Data Policy*, FACEBOOK, http://www.facebook.com/policy.php (last modified Sept. 29, 2016); *Privacy Policy*, GOOGLE, http://www.google.com/policies/privacy/#infocollect (last modified Mar. 1, 2017).

118. Ferguson, *supra* note 4, at 376.

119. *See* Epstein, *supra* note 111, at 1200 (discussing the limits of a reasonable expectation of privacy with third parties).

120. Ferguson, *supra* note 4, at 331.

121. Douglas Page, *Crime Fighting's Next Big Deal*, OFFICER.COM (Sept. 4, 2012), http://www.officer.com/article/10773317/crime-fightings-next-big-deal; *see also* Press Release, Microsoft, New York City Police Department and Microsoft Partner to Bring Real-Time Crime Prevention and Counterterrorism Technology Solution to Global Law Enforcement Agencies (Aug. 8, 2012), http://news.microsoft.com/2012/08/08/new-york-city-police-department-and-microsoft-partner-to-bring-real-time-crime-prevention-and-counterterrorism-

with Internet posts that have been abandoned or left in the public domain can reveal a broad swath of information about an individual that prior to the harnessing of big data would be unascertainable.[122]

Returning to the earlier example of the young man lingering outside of the convenience store at 9:24 p.m., the officer in that hypothetical situation was able to make a presumptive positive identity match based on the young man's characteristics by searching a criminal record database. In New York City, when the officer collects information from her own observations (e.g., lingering, putting hands in pockets, fidgeting, etc.) and augments her impressions with data from the DAS database and Internet content published by the suspect, she may justify conducting an investigatory stop of the suspect because she can articulate a particularized and individualized set of facts warranting a reasonable suspicion.[123] As suggested by Professor Ferguson, "in a criminal investigation, the inferences of suspicion are easy to develop and, against a low legal threshold, easy to meet."[124] However, this is not *Minority Report* or a way to police malevolent thoughts. The officer's initial subjective impressions of what she believes to be suspicious behavior initiates the law enforcement actions of querying the available databases, scanning available Internet posts with web crawler software, and ultimately making an investigatory stop.

C. Benefits and Disadvantages of Big Data Analytics and Social Media

While the subjective nature of what constitutes suspicious behavior is problematic because of the potential for bias, prejudice, and interest,[125] the law enforcement effort in this instance focuses on deterring a crime that is already in progress but that has not yet reached fruition (as opposed to investigating a completed crime). Bias and prejudice informing the impressions of a law enforcement officer present a risk to the public irrespective of whether big data is employed to pursue initial inklings of suspicious behavior. In addition, reasonable suspicion itself presents inherent logical weaknesses when coupled with big data analytics to perform police work.[126]

Reasonable suspicion supported with big data creates a self-reinforcing feedback loop that is indicative of circular reasoning.[127] The self-reinforcing feedback loop begins with an observation by law enforcement of a person that appears suspicious.[128] An officer with access to databases of criminal activity, criminal history, and Internet crawlers then locates information that supports

technology-solution-to-global-law-enforcement-agencies/.

122. *See generally* Solove, *supra* note 63, at 1186 (arguing it is the aggregation of the individual datum into an interconnected mass that makes information more powerful).

123. Ferguson, *supra* note 4, at 379.

124. *Id.* at 380.

125. Jason Marsh, *Can We Reduce Bias in Criminal Justice?*, GREATER GOOD SCI. CTR. (Apr. 28, 2015), http://greatergood.berkeley.edu/article/item/can_we_reduce_bias_in_criminal_justice.

126. Ferguson, *supra* note 4, at 387.

127. ROBERT D. COLEMAN, WHAT IS CIRCULAR REASONING? (2006), http://www.numeraire.com/download/WhatIsCircularReasoning.pdf.

128. *See generally id.* (describing the circular reasoning fallacy).

the hypothesis that the person of interest is behaving suspiciously, fits the profile of a criminal, or, if a positive identification has been made, has a criminal history. Confirmation bias then may color the picture drawn by the data gleaned from the database query because the officer may tend to search for information that validates and verifies the hypothesis while disregarding information that would render it null.[129] With each piece of inculpating information added, the initial hypothesis that the suspect was behaving suspiciously is reinforced in a swirling circular reasoning.[130] As each factor contributes to reinforcing the reasonable suspicion necessary to conduct an investigatory stop, the mounting bits of information shield the investigatory stop from legal scrutiny.

Another potential problem is the level of investment required to train law enforcement officers in the street-level application of big data analytics and descriptive statistics. Similarly, data sets may not be sufficiently large in the case of small towns to be able to conduct *Terry* stops based on statistical inferences. Any inference incorporating data from metropolitan areas and applying insights to lower-population areas may result in greater numbers of false positives in the smaller community, despite a reduction in the potential for bias in aggregated, anonymized data sets.

Courts consider the totality of the circumstances when analyzing reasonable suspicion.[131] Whether the initial impression of suspicious behavior was reasonable determines the validity of the subsequent actions by the law enforcement officer.[132] Courts would analyze whether the behavior was reasonably suspicious to warrant the officer's use of information published to the public or contained in a database.[133] If the behavior and the information together provided a sufficient basis for the officer having a reasonable suspicion, then the investigatory stop was justified.[134]

The nature of the totality-of-the-circumstances test creates additional problems for the doctrine of reasonable suspicion in that the quantity of factors can supersede the quality of factors and the information gleaned from data sources can be contorted to fit the "individualized and particularized" requirements of *Terry v. Ohio*.[135] It follows then that when quantity of information is a byproduct of big data analysis, a police officer can cull sufficient amounts of information from database queries to express a coherent suspicion under a totality-of-the-circumstances analysis based on characteristics rather than behavior, thus undermining the Fourth Amendment's safeguard.[136] Second, the language of *Terry v. Ohio* is very

129. *Confirmation Bias*, INVESTOPEDIA, http://www.investopedia.com/terms/c/confirmation-bias.asp (last visited Mar. 11, 2017).

130. *Id.*

131. Ferguson, *supra* note 4, at 387.

132. *See* Terry v. Ohio, 392 U.S. 1, 27 (1968) (discussing the analysis of a justified investigatory stop).

133. *Id.*

134. *Id.*

135. *See generally* Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461, 463–64 (2015) (articulating the reality that courts have "come to expect" very detailed accounts from law enforcement to give rise to a reasonable suspicion).

136. *Id.*

general and does not differentiate between crimes and criminals.[137] In Professor Ferguson's "small data" world, the distinction was immaterial for law enforcement since presumably the observations of criminals and criminal activity would be identical.[138]

In the big data reality, the law enforcement officer must tie the suspicious behavior to the suspected crime.[139] Having an assortment of data points unconnected to the crime suspected is unreasonable suspicion. Someone who has been arrested for domestic battery three times does not give rise to a suspicion that he is selling drugs if observed loitering outside a convenience store. However, three prior arrests for possession with intent to distribute may be particularly persuasive in a reasonable suspicion inquiry.

Big data gleaned from web pages, publicly viewable Internet posts, criminal record databases, and other sources, while fraught with potential for bias and prejudice, offers society several law enforcement paradigm-shifting benefits.[140] First, more data correlates to greater accuracy in criminal investigations by focusing the investigatory scope, in turn saving taxpayers significant sums of money.[141] As Professor Baradaran points out, law enforcement has considerable freedom in discerning what constitutes suspicious behavior, as well as deciding where to patrol and whom to observe or investigate further.[142] Everyone makes instant judgments about the people around them, and law enforcement officers are subject to the same human errors of false attribution, bias, prejudice, and self-interest as any other person.[143] Substituting data from databases for some of the cognitive judgments by law enforcement enables reasonable suspicion to be bolstered by indiscriminating software code while also narrowing the persons of interest in criminal investigations.[144] A byproduct of increased accuracy is the potential to foster better relationships with the communities that feel marginalized by police.[145] By reducing the necessity to confirm or deny a reasonable suspicion by effecting an investigatory stop, big data verification can empower police to simultaneously focus on building relationships with citizens of the community and to protect themselves from potential violence during routine stops.[146]

---

137. Ferguson, *supra* note 4, at 387.

138. *Id.* at 388.

139. *Id.*

140. *Id.*

141. *See generally* King, *supra* note 14 (describing how using analytics has made police work more efficient).

142. *See generally* Baradaran, *supra* note 2 (discussing need to revise the role of the balancing test in Fourth Amendment jurisprudence because it preserves social inequities and endows law enforcement without erecting clear limitations).

143. *See* L. Song Richardson, *Cognitive Bias, Police Character, and the Fourth Amendment*, 44 ARIZ. ST. L.J. 267, 270–73 (2012) (articulating that implicit biases subtly affect judgment and cognitive decision making).

144. Ferguson, *supra* note 4, at 390.

145. *Id.* at 391.

146. *See* David Rudovsky, *Law Enforcement by Stereotypes and Serendipity: Racial Profiling and Stops and Searches Without Cause*, 3 U. PA. J. CONST. L. 296, 334 (2001) (observing and providing an example of friction between marginalized communities and law enforcement generated during investigatory stops).

A corollary to increased accuracy in police work is reduced prosecution of innocent citizens, increased transparency, and greater individual accountability.[147] The potential to become embroiled in a self-reinforcing loop exists, but courts may consider suspicion-reducing factors in a totality-of-the-circumstances analysis and place the same weight on them as they do the inculpating facts.[148] Under this framework, big data in policing also increases transparency through accountability.[149] When an officer observes what she considers reasonably suspicious behavior, her subsequent data queries, Internet data gathering, and search results can all be accessed after the fact and become discoverable in ensuing litigation.[150] Thus, rather than compel an officer to justify her reasonable suspicion after the fact using her memory, the documented record of her big data query can be produced for inspection and cross-examined at trial, resulting in greater transparency and accountability in the police process.

D. The ABA's Reply—Law Enforcement Access to Third-Party Records

The ABA first published recommended guidelines for law enforcement's use of electronic surveillance in 1971 as part of its Standards for Criminal Justice.[151] In 2013, the ABA's most recent recommendations came off the press with the publication of the Criminal Justice Standards on Law Enforcement Access to Third Party Records (LEATPR Standards).[152] While the ABA LEATPR Standards do not carry the force of law, they are persuasive authority in courts with respect to the government's desire to access third-party records.[153] The LEATPR Standards specifically limit the ambit of their operation to law enforcement's access of third-party records for use in a criminal investigation, while refraining from commenting on records impacting national security, preventing terrorism, or prosecuting criminals.[154] In articulating the blueprint for approaching government access to third-party records, the drafters of the LEATPR Standards recognized that the Court disallowed a continuing reasonable expectation of privacy to information voluntarily given to third parties.[155] Nonetheless, they sought to categorize information according to what level of privacy the ABA believes attaches to varieties of information held by third parties.[156]

147. *See* Joshua A.T. Fairfield & Erik Luna, *Digital Innocence*, 99 CORNELL L. REV. 981, 1031 (2014) (arguing that law enforcement has a positive duty to provide any positive information or favorable exculpatory evidence when the government uses big data analytics to investigate and develop a case).

148. *See* COLEMAN, *supra* note 127 (describing circular reasoning); *see also* Ferguson, *supra* note 4, at 392.

149. Ferguson, *supra* note 4, at 393.

150. *Id.*

151. STANDARDS FOR CRIMINAL JUSTICE: ELEC. SURVEILLANCE (AM. BAR ASS'N 1971).

152. STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS (AM. BAR ASS'N, amended 2013).

153. *Persuasive Authority*, BLACK'S LAW DICTIONARY (9th ed. 2009).

154. STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS at 5.

155. *Id.* at 6–10.

156. *Id.*

Privacy attaching to information under the LEATPR Standards is characterized in a hierarchical framework wherein information the ABA believes has the greatest expectation of remaining private is at the top and information that has no privacy expectation is at the bottom.[157] Specifically, the ABA characterizes information as "highly private, moderately private, minimally private, [and] not private."[158] In order to determine what characterization applies to the information held by the third party, the ABA articulates four factors for analysis.[159]

The first factor in characterizing the privacy level attaching to information is the reasonable necessity of the divulgence to the third party for the individual "to participate meaningfully in society or in commerce, or is socially beneficial . . . ."[160] If a person must make an information disclosure in order to "participate meaningfully" in society, then an expectation of privacy to that information is more reasonable, as in the case of financial records.[161] The second factor in the analysis is whether the information is of an inherently personal nature.[162] In deciding whether information is personal, the probability of embarrassment or stigma resulting from disclosure of the information is important, as in the case of text messages about a medical diagnosis.[163] Third, the ABA considers the information's accessibility by third parties other than the initial recipient of the information transmission.[164] A person who publishes information online that is easily accessed by others holds a lower expectation of privacy than information that is less easily accessed by others, as in the case of Facebook posts versus geographical information from a telecommunications company.[165] The final factor in determining the level of privacy that attaches to information under the LEATPR Standards is whether statutory or case law establishes protections for the information, as in the case of health information.[166] Importantly, state laws often control criminal investigations, allowing variations in the latitude provided to law enforcement and prosecutors who seek to obtain third-party records.[167]

Once the information held by a third party has been characterized using the four-factor analysis, a corresponding level of protection attaches to the information.[168] The ABA recommends that "highly private information should be highly protected . . . , moderately private information should be moderately protected . . . , minimally private information should be minimally protected . . . , and . . . information that is not private should be

157. *Id.* at 10.
158. *Id.*
159. *Id.* Standard 25-4.1, at 19–20.
160. *Id.* Standard 25-4.1(a), at 20.
161. *Id.* Standard 25-4.1(a) cmt. at 57–65.
162. *Id.* Standard 25-4.1(b), at 20.
163. *Id.* Standard 25-4.1(b) cmt. at 65–77.
164. *Id.* Standard 25-4.1(c), at 20.
165. *Id.* Standard 25-4.1(c) cmt. at 77–80 (describing the ABA's rationale that information desired to be private is less likely to be disclosed to multiple third parties as compared to less-private information).
166. *Id.* Standard 25-4.1(d), at 20.
167. *Id.* Standard 25-4.1(d) cmt. at 80–90.
168. *Id.* Standard 25-4.2(a), at 20.

unprotected."[169] The protection for the information neatly maps onto Fourth Amendment jurisprudence insofar as highly protected information requires probable cause and a warrant, moderately protected information requires reasonable suspicion, minimally protected information can be accessed on a showing of relevance, and unprotected information is already readily accessible.[170] For use of information void of personally identifying information, the ABA recommends that law enforcement agencies have full access to third-party data stores of these anonymous records.[171]

Law enforcement would still be able to access the underlying personally identifiable information if the threshold for obtaining that particular species of protected information were reached (e.g., reasonable suspicion for moderately protected information, etc.).[172] Thus, in the context of leveraging massive data sets to investigate a crime, the LEATPR Standards neatly categorize information into classes of data with corresponding protections.[173] However, they do not adequately cover Internet communications on social media sites, blogs, or other Internet data sources used in data analysis.[174] Consider the example of the young man outside the convenience store once more. In that example, reasonable suspicion was manufactured from aggregated records and police databases. Under the LEATPR Standards, this hypothetical reasonable suspicion would enable a law enforcement officer to access moderately private information specific to the presumptive identity of the individual.[175] Once that private information specific to the individual can be accessed, an array of personal information may be laid bare that otherwise would be protected.[176]

Professor Ferguson hypothesizes that the relationship described above between big data and Fourth Amendment requirements for searches and seizures undermines the spirit of the law's protections.[177] When legal requirements for searches and seizures are being fulfilled by a brief nod to the letter of the law by using big data to generate a reasonable suspicion, the protections guaranteed by that amendment are eroded.[178] To be sure, the ABA's LEATPR Standards provide a moderate path between the genuine need law enforcement has in accessing third-party records and the equally genuine expectation of privacy in one's personal affairs. Nevertheless, the Standards do not adequately consider the panoply of complications to privacy considerations when communications are transmitted via the Internet to a reasonably limited audience.

---

169. *Id.*
170. *Id.* at 10.
171. *Id.* Standard 25-5.6(a), at 23.
172. *Id.* Standard 25-5.6(b), at 23.
173. *Id.* Standard 25-4.2(a), at 20; Ferguson, *supra* note 35.
174. Ferguson, *supra* note 35.
175. STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS at 10.
176. *See* Ferguson, *supra* note 35, at 831 ("The value in 'third party records' is information—masses of revealing information.").
177. *See id.* at 834 ("[T]he amount and interconnectedness of the available data weakens legal standards like 'relevance,' 'reasonable suspicion,' and 'probable cause' . . . .").
178. *See id.* at 835 (discussing the "distortion" of the legal standards under the Fourth Amendment when considered in light of law enforcement's use of big data).

IV. RECOMMENDATION

Congress should enact a law that requires courts to reinforce Fourth Amendment protections for Internet posts. The law should require courts to weigh an Internet post's privacy expectation under the objective ABA's LEATPR Standards, together with certain plus factors, to determine whether it is subject to data aggregation and web indexing.[179] A plus factor here may be proactive actions to keep transmitted Internet communications private.[180] Additionally, the law should require courts to conduct an analysis for each data query and Internet-post pull that the court determines was predominately supported by correlations between characteristics of persons who have committed a crime and the person observed. The courts determine if the individualized observations together with the descriptive statistics are sufficiently particularized to give rise to reasonable suspicion. This would aid in precluding extensive digital investigations based on descriptive statistics.[181] The need for such a test is made manifest by the myriad complexities of privacy protections that may arise in cases subject to varying standards at the state level.[182] While the ABA's LEATPR Standards articulate a novel approach to classifying information, the recommendations focus on discrete interactions with data that provide law enforcement a blanket authorization to delve deeper into personally identifiable information once some level of relevance or reasonable suspicion can be articulated.[183] Moreover, the Electronic Communications Privacy Act of 1986 (ECPA) protects a limited class of personal information from law enforcement investigations—namely, e-mail and files stored on a remote server—and does not reach transmissions to third parties outside of these protected classes.[184]

Law enforcement has a valid and justifiable need for information that can be accessed online.[185] Yet, when a police officer sees a person and conducts a search of their personal digital identity based on descriptive statistics resulting from analytics as though they were running a license plate, Fourth Amendment protections are implicated. This is because a digital search of personal information is analogous to an investigatory stop of the person and is unlike

179. *See* Kovacic et al., *supra* note 18 (articulating the use of plus factors as circumstantial evidence in antitrust law).

180. *See generally* Horton v. California, 496 U.S. 128, 133 (1990) (reasoning that individuals have an interest in privacy, but when evidence is left in plain view no proactive actions have been taken to keep the item private).

181. *See generally* Ferguson, *supra* note 4 (discussing how the Fourth Amendment is applied in the law enforcement context and the requirement of reasonable suspicion).

182. *See generally* Alicia Shelton, *A Reasonable Expectation of Privacy Online: "Do Not Track" Legislation*, 45 U. BALT. L.F. 35, 48, 49 (2014) (arguing that the legislature is "the proper body to establish protection of individuals' privacy").

183. *See* Ferguson, *supra* note 35, at 834 (hypothesizing that the LEATPR Standards encounter problems when considered in conjunction with data analytics); *see also* STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS (AM. BAR ASS'N, amended 2013) (classifying data into levels of privacy expectations).

184. Deirdre K. Mulligan, *Reasonable Expectations in Electronic Communications: A Critical Perspective on the Electronic Communications Privacy Act*, 72 GEO. WASH. L. REV. 1557, 1557–59 (2004).

185. *See* Ferguson, *supra* note 4, at 388 ("[L]aw enforcement officials see the potential of these tools to reduce crime.").

obtaining registration information from a state's department of motor vehicles.[186] While the Supreme Court has announced the principle that "what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection," it also recognizes "what he seeks to preserve as private, even in an area accessible to the public, may be constitutionally protected."[187] Thus, Congress should require that a law enforcement officer's hunch or observation giving rise to a *Terry* stop be justified by something more than data analytics and an observed characteristic that correlates to the analytical output. This is necessary to preserve the strength of the constitutional protection.[188] This is not to say, though, that characteristics are disposed of entirely. For example, in the context of organized crime, tattoos matching known gang affiliations, or clothing and colors associated with gang activity, are some characteristics that may give rise to a reasonable suspicion to conduct a database search or integrate data analytics.[189] However, immutable characteristics such as height, weight, and race should not justify a reasonable suspicion. This process should be iterative and occur for every data query.[190] Thus, if a web crawler program returns several social media posts in the public domain belonging to the person who has been presumptively identified, pursuing the investigation further should require reasonable suspicion based on the results of the query or observed behavior of the person.[191] Reasonable suspicion adjudicated under this congressional mandate allows sufficient leeway for law enforcement to use big data while reinforcing the Fourth Amendment's protection against investigatory stops supported solely by premonitions, feelings, or hunches.[192]

Similarly, an Internet search of a person's content on social media must be weighed in consideration of the person's expectation of privacy that attaches to that species of information under the ABA LEATPR's Standards, together with plus factors that indicate a desire to prevent the dissemination of that information.[193] For example, a photo posted on a blog or social media website generally viewable by the public that depicted a person using narcotics or preparing them for distribution, and included a recent time stamp, would

186. *See generally* U.S. CONST. amend. IV (announcing the right to be secure in a person's papers and effects).

187. Katz v. United States, 389 U.S. 347, 351 (1967).

188. *See* Ferguson, *supra* note 4, at 331 ("This new reality simultaneously undermines the protection that reasonable suspicion provides . . . and potentially transforms reasonable suspicion into a means of justifying those same stops.").

189. *See* Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY. L.J. 259, 266 (2012) (describing how predictive analytics relates to gang activity).

190. *See* Ferguson, *supra* note 4, at 405 ("[C]ourts might require more information to satisfy the reasonable suspicion standard.").

191. *See id.* (describing one way to improve the use of big data in law enforcement, which is to make doctrinal changes that result in a reinforced reasonable suspicion standard).

192. *Id.*

193. *See* STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS (AM. BAR ASS'N, amended 2013) (classifying data into levels of privacy expectations); Kovacic et al., *supra* note 18 (articulating the use of plus factors as circumstantial evidence in antitrust law); Horton v. California, 496 U.S. 128, 133 (1990) (reasoning that individuals have an interest in privacy, but when evidence is left in plain view no proactive actions have been taken to keep the item private).

justify a reasonable suspicion that the person may be selling narcotics.[194] Thus, under the balancing approach just described, this person would enjoy no Fourth Amendment protections.[195] Conversely, a person who posts content involving narcotic use and personal health information to a social media account that precludes the general public from viewing the content of the page and even restricts the post's viewable audience to an even more limited audience, say only to family, would have a substantially greater expectation of privacy.[196] This balancing approach would encourage courts to consider the proactive actions the person took to maintain the privacy and to limit the dissemination of the communication as a plus factor favoring a higher degree of privacy protection.[197] This proposed law meets the Court's concern in *Katz* that "what [a person] seeks to preserve as private, even in an area accessible to the public, may be constitutionally protected."[198]

This proposed law incorporates the ABA's LEATPR Standards in order to mitigate the potential for invasive searches of American citizens by law enforcement while simultaneously allowing for the broad, but judicious use of massive data set analytics by police to combat crime and make investigatory stops based on descriptive statistics and/or social media posts without any privacy protections.[199] By maintaining an iterative requirement for law enforcement to articulate a reasonable suspicion, this test would accommodate a law enforcement officer's initial observations and analytical outputs to perform initial Internet searches analogous to digital investigatory stops, but limit their number to curtail the potential to manufacture a continuing reasonable suspicion.[200] Similarly, by taking into account the underlying information posted online together with plus factors, such as actions to limit distribution, this law would uphold the intention of the Fourth Amendment's Framers to provide American citizens freedom from unwarranted government intrusion into their private lives, as well as encourage transparency and accountability in law enforcement.[201]

V. CONCLUSION

Control over one's personal affairs, free from government intrusion, was the motivating animus behind the Bill of Rights.[202] The Fourth Amendment's

194. *See* United States v. Raney, 342 F.3d 551, 558–59 (7th Cir. 2003) (illustrating police search and seizure of photo when consent to search is given).

195. *Id.*

196. *See* LAW ENF'T ACCESS TO THIRD PARTY RECORDS Standard 25-4.1 cmt. at 55–57 (articulating different examples such as personal health information that yield greater degrees of privacy expectations).

197. *Id.* Standard 25-4.1(d), at 19–20.

198. Katz v. United States, 389 U.S. 347, 351 (1967).

199. *See* Ferguson, *supra* note 4, at 379.

200. *See id.* ("With each level of search, officers can access additional individualized and particularized facts that, when viewed within the totality of circumstances, help justify the officer's stop of a suspect. The effect is that additional personalized information encourages a finding of reasonable suspicion.").

201. *See generally* Thomas K. Clancy, *The Framers' Intent: John Adams, His Era, and the Fourth Amendment*, 86 IND. L.J. 979 (2011) (providing an in-depth treatment of the Fourth Amendment, its intent, and the relationship between citizens and the government as considered at the time of the Bill of Rights).

202. *See* Joseph Postell, *Securing Liberty: The Purpose and Importance of the Bill of Rights*, HERITAGE

protective scope sought to insulate the home, the body, and the personal affairs of citizens from government surveillance and control.[203] As computing power has increased and the role of the Internet has expanded, the protections attaching to information have become dubious.[204] Where reasonable suspicion was once the result of direct observations of behavior, law enforcement's initial perceptions may now be augmented by statistical outputs from data analytics that could give rise to a legally valid justification to conduct an investigatory stop.[205]

In a reality where data is ubiquitous and crime is a clear threat, law enforcement's use of analytics to focus patrol efforts and direct attention to communities hardest hit by crime is a boon to taxpayers and police officers alike.[206] But, the use of data analytics, web crawlers, and other technology must be conducted in accord with the Fourth Amendment's requirements for criminal investigations.[207] The effect of using an unreasonable suspicion—such as observing a young man leaning against a convenience store—to justify further investigation into a person's digital life should amount to an unlawful search or seizure of that person's personal effects. To be lawful, a search or seizure must be justified by a reasonable suspicion undergirded by something more than one discrete observation. This Note has proposed that Congress pass a law that creates a balancing test courts should use to determine whether an investigation into a person's digital life was reasonable; the purpose of this law would be to expand the scope of the Fourth Amendment protections. The proposed test weighs the expectation of privacy that the information could reasonably be expected to maintain (per the ABA's LEATPR Standards) against the reasonable suspicion articulated by the officer who conducted the search.[208] The test seeks a middle ground in order to mitigate the potential for invasive searches by law enforcement while simultaneously allowing for the broad but judicious use of massive data set analytics by police to combat crime and make investigatory stops based on data analytics, statistical inferences, and social media postings.

In the end, the proper role and function of data analytics and the appropriate level of privacy the law provides to social communications will be defined by statutory law and shaped by judicial decision.[209] However, there

---

FOUND. (Dec. 14, 2007), http://www.heritage.org/research/reports/2007/12/securing-liberty-the-purpose-and-importance-of-the-bill-of-rights (detailing the Bill of Rights' ratification and underlying purpose).

203. *See* Clancy, *supra* note 201, at 1060 ("Adams identified four objects as protected: people, houses, papers, and possessions.").

204. *See* Ferguson, *supra* note 4, at 331 (discussing the role of reasonable suspicion standard relative to data analytics).

205. *Id.*

206. *See generally* King, *supra* note 14 (describing how using analytics has made police work more efficient).

207. *See* Horton v. California, 496 U.S. 128, 133 (1990) (reasoning that individuals have an interest in privacy, but when evidence is left in plain view no proactive actions have been taken to keep the item private); Katz v. United States, 389 U.S. 347, 351 (1967).

208. *See* STANDARDS FOR CRIMINAL JUSTICE: LAW ENF'T ACCESS TO THIRD PARTY RECORDS (AM. BAR ASS'N, amended 2013); Ferguson, *supra* note 4, at 405 (describing one way to improve the use of big data in law enforcement is to make doctrinal changes that result in a reinforced reasonable suspicion standard).

209. *See generally* Shelton, *supra* note 182 (arguing that the legislature is "the proper body to establish

are limits to law enforcement's use of data analytics. This proposed balancing test upholds the Framers' intention behind the Fourth Amendment to ensure individual freedom of expression and communication and to protect individuals from unwarranted government intrusion while also facilitating law enforcement's efforts to protect local communities from crime.[210]

protection of individuals' privacy").

210. Clancy, *supra* note 201.