UNIVERSITY *of* PENNSYLVANIA
LAW REVIEW

Founded 1852

Formerly
AMERICAN LAW REGISTER

VOL. 163 JANUARY 2015 NO. 2

ARTICLE

BIG DATA AND PREDICTIVE REASONABLE SUSPICION

ANDREW GUTHRIE FERGUSON[†]

[†] Associate Professor of Law, David A. Clarke School of Law, University of the District of Columbia. With appreciation to Professors Jack Chin, Daniel Filler, Orin Kerr, Richard Leo, Daniel Medwed, Jenny Roberts, Josephine Ross, David Rudovsky, Laurent Sacharoff, Christopher Slobogin, and Joe Tulman for their comments and suggestions. I would like to dedicate this Article to the late Professor Andy Taslitz, who encouraged me to write it and who was an inspiring and supportive friend to many scholars.

> *[F]rom a legal point of view there is nothing inherently unattainable about a prediction of future criminal conduct.*[1]

> *Electronic databases form the nervous system of contemporary criminal justice operations. In recent years, their breadth and influence have dramatically expanded. . . . The risk of error stemming from these databases is not slim. . . . Inaccuracies in expansive, interconnected collections of electronic information raise grave concerns for individual liberty.*[2]

INTRODUCTION

The Fourth Amendment requires "reasonable suspicion" to stop a suspect.[3] As a general matter, police officers develop this suspicion based on information they know or activities they observe. Suspicion is individualized to a particular person at a particular place.[4] Most reasonable suspicion cases involve police confronting unknown suspects engaged in observable suspicious activities.[5] Essentially, the reasonable suspicion doctrine is based on "small data"—discrete facts, limited information, and little knowledge about the suspect.[6]

But what happens if this small data suspicion is replaced by "big data" suspicion?[7] What if police can "know" personal information about the suspect by searching vast networked information sources? The rise of big data technologies offers a challenge to the traditional paradigm of Fourth Amendment law. With little effort, officers can now identify most unknown

[1] Schall v. Martin, 467 U.S. 253, 278 (1984).

[2] Herring v. United States, 555 U.S. 135, 155 (2009) (Ginsburg, J., dissenting).

[3] *See* Terry v. Ohio, 392 U.S. 1, 27 (1968).

[4] *See, e.g.*, United States v. Arvizu, 534 U.S. 266, 273 (2002) ("[W]e have said repeatedly that [courts] must look at the 'totality of the circumstances' of each case to see whether the detaining officer has a 'particularized and objective basis' for suspecting legal wrongdoing.").

[5] *See infra* Part I.

[6] "Small data," like "big data," has no set definition. Generally, small data is thought of as solving discrete questions with limited and structured data, and the data are generally controlled by one institution. *See generally* JULES J. BERMAN, PRINCIPLES OF BIG DATA: PREPARING, SHARING, AND ANALYZING COMPLEX INFORMATION 1-2 (2013).

[7] *See generally* Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1920-21 (2013) ("'Big Data' is shorthand for the combination of a technology and a process. The technology is a configuration of information-processing hardware capable of sifting, sorting, and interrogating vast quantities of data in very short times. The process involves mining the data for patterns, distilling the patterns into predictive analytics, and applying the analytics to new data. Together, the technology and the process comprise a technique for converting data flows into a particular, highly data-intensive type of knowledge."); Steve Lohr, *Amid the Flood, A Catchphrase Is Born*, N.Y. TIMES, Aug. 12, 2012, at BU3 [hereinafter Lohr, *Amid the Flood*] ("Big Data is a shorthand label that typically means applying the tools of artificial intelligence, like machine learning, to vast new troves of data beyond that captured in standard databases.").

suspects, not through their observations, but by accessing a web of information containing extensive personal data about suspects.[8] New data sources, including law enforcement databases, third-party records, and predictive analytics, combined with biometric or facial recognition software, allow officers access to information with just a few search queries.[9] At some point, inferences from this personal data (independent of the observation) may become sufficiently individualized and predictive to justify the seizure of a suspect. The question this Article poses is whether a Fourth Amendment stop can be predicated on the aggregation of specific and individualized, but otherwise noncriminal, factors.

For example, suppose police are investigating a series of robberies in a particular neighborhood. Arrest photos from a computerized database are uploaded in patrol cars. Facial recognition software scans people on the street.[10] Suddenly there is a match—police recognize a known robber in the targeted neighborhood. The suspect's personal information scrolls across the patrol car's computer screen—prior robbery arrests, prior robbery convictions, and a list of criminal associates also involved in robberies.[11] The officer then searches additional sources of third-party data, including the suspect's GPS location information for the last six hours or license plate records which tie the suspect to pawn shop trades close in time to prior robberies.[12] The police now have particularized, individualized suspicion about a man who is not doing anything overtly criminal. Or perhaps predictive software has already identified the man as a potential reoffender for this particular type of crime.[13] Or perhaps software has flagged the suspect's social media comments or other Internet postings that suggest planned criminal or gang

[8] *See infra* Part II.

[9] *See infra* Part II.

[10] *See infra* Part II; *see also Cop Car with Built-In Face Recognition and Predictive Policing Wins UK Award*, PRIVACYSOS.ORG (Apr. 4, 2013, 4:10 PM), http://privacysos.org/node/1016, *archived at* http://perma.cc/Y7BA-NTV2 (highlighting an example of technological advances in policing).

[11] All of this information is available through a National Crime Information Center (NCIC) search. *See National Crime Information Center*, FBI, http://www.fbi.gov/about-us/cjis/ncic (last visited Nov. 7, 2014), *archived at* http://perma.cc/P3CG-M5HF (explaining resources for finding information about criminals). This information is further available through police computers accessible in police cars and in police stations. *See National Crime Information Center Celebrates 40th Birthday*, GOV'T TECH. (Jan. 22, 2007), http://www.govtech.com/gt/articles/103437, *archived at* http://perma.cc/PDL7-JKS6 (discussing how NCIC records have helped law enforcement).

[12] *See infra* Part II.

[13] Local jurisdictions sometimes create their own "most wanted" lists of locally identified criminals. *See, e.g.*, *Most Wanted*, L.A. POLICE DEP'T, http://www.lapdonline.org/most_wanted (last visited Nov. 7, 2014), *archived at* http://perma.cc/9P5R-KZRG (showing a local jurisdiction's most wanted list).

activity.[14] Can this aggregation of individualized information be sufficient to justify interfering with a person's constitutional liberty?

This Article traces the consequences of a shift from "small data" reasonable suspicion, focused on specific, observable actions of unknown suspects, to a "big data" reality of an interconnected, information rich world of known suspects. With more specific information, police officers on the streets may have a stronger predictive sense about the likelihood that they are observing criminal activity.[15] This evolution, however, only hints at the promise of big data policing. The next phase will use existing predictive analytics to target suspects without any firsthand observation of criminal activity, relying instead on the accumulation of various data points.[16] Unknown suspects will become known to police because of the data left behind.[17] Software will use pattern-matching techniques[18] to identify individuals by sorting through information about millions of people contained in networked databases. This new reality simultaneously undermines the protection that reasonable suspicion provides against police stops and potentially transforms reasonable suspicion into a means of justifying those same stops.

This Article seeks to offer three contributions to the development of Fourth Amendment theory. First, it demonstrates that reasonable suspicion—as a small data doctrine—may become practically irrelevant in an era of big

[14] *See, e.g.*, Heather Kelly, *Police Embrace Social Media as Crime-Fighting Tool*, CNN (Aug. 30, 2012, 5:23 PM), http://www.cnn.com/2012/08/30/tech/social-media/fighting-crime-social-media, *archived at* http://perma.cc/D2LC-DEH8 (detailing ways police officers use social media to catch or thwart criminals).

[15] While this may protect some individuals who are not likely to be involved in criminal activity, it may also create additional burdens on those who are predicted to be involved in criminal activity. *See infra* Part IV.

[16] *See infra* Part II.

[17] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716 (2010) (highlighting the difficulty of protecting the privacy of data subjects by anonymizing data); Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1877-78 (2011) (discussing identification of individuals from personally identifiable information found from data sources); Rebecca J. Rosen, *Stanford Researchers: It Is Trivially Easy to Match Metadata to Real People*, ATLANTIC (Dec. 24, 2013, 1:50 PM), http://www.theatlantic.com/technology/archive/2013/12/stanford-researchers-it-is-trivially-easy-to-match-metadata-to-real-people/282642/, *archived at* http://perma.cc/QFK5-6JUC (explaining the ease with which metadata can be matched with specific individuals).

[18] *See* Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 4 (2005) ("Data mining's computerized sifting of personal characteristics and behaviors (sometimes called 'pattern matching') is a more thorough, regular, and extensive version of criminal profiling, which has become both more widespread and more controversial in recent years."); Gareth Cook, *Software Helps Police Draw Crime Links*, BOS. GLOBE, July 17, 2003, at A1 (discussing how law enforcement officers are using databases as research tools).

data policing. Second, it examines the distortions of big data on police observation, investigation, and prediction, concluding that big data information will impact all major aspects of traditional policing. Third, it seeks to offer a solution to potential problems using the insights and value of big data itself to strengthen the existing reasonable suspicion standard.

Part I of this Article examines the development of Fourth Amendment law on reasonable suspicion. Much of this case law involves "unknown" suspects, such as when a police officer sees an individual on the street but does not know his or her identity. In these cases, reasonable suspicion necessarily derives from the suspect's observable actions. Most Fourth Amendment cases involving police–citizen encounters are of this "stranger" variety.[19] Thus, the reasonable suspicion test, as it evolved, required the police officer to articulate individualized, particularized suspicion to distinguish a stranger's suspicious actions from non-suspicious actions.[20] The resulting doctrine, created around actions, not individuals, makes sense within the context it arose (as presumably most officers would not know all of the potential criminals in their patrol areas).[21] The resulting reasonable suspicion test, however, becomes significantly distorted when officers have access to more individualized or predictive information about a suspect.

Part II of this Article addresses the rise of "big data" in criminal law enforcement. Law enforcement organizations are working to grow the scope, sophistication, and detail of their databases.[22] Agencies and their officers may now search national databases and gain instant access to the information.[23] Indeed, "data" is the new watchword in many smart-policing districts.[24]

[19] *See infra* Part I.

[20] *See* William J. Mertens, *The Fourth Amendment and the Control of Police Discretion*, 17 U. MICH. J.L. REFORM 551, 594-95 (1984) ("[T]he police must be able to justify singling out from the rest of humanity (or at least from the rest of the people in the general area) the particular individual whom they have stopped as somehow meriting this special attention.").

[21] This assumption is certainly true in large urban police districts, although it may hold less true for small towns or rural areas. As will be discussed later, "big data" in some ways turns big city policing into old-fashioned, small-town policing, with the benefits and drawbacks that come from that scale of police surveillance.

[22] *See infra* Part II.

[23] *See infra* Part II.

[24] *See* Nina Cope, *Intelligence Led Policing or Policing Led Intelligence?*, 44 BRIT. J. CRIMINOLOGY 188, 191 (2004) (discussing an operational structure for the organization of intelligence processes in police forces); Stephen Baxter, *Modest Gains in First Six Months of Santa Cruz's Predictive Police Program*, SANTA CRUZ SENTINEL (Feb. 26, 2012, 4:59 PM), http://www.santacruzsentinel.com/rss/ci_20050377, *archived at* http://perma.cc/KPM5-K634 (reporting on the success of a data algorithm used by the Santa Cruz Police Department); Carrie Kahn, *At LAPD, Predicting Crimes Before They Happen*, NPR (Nov. 26, 2011, 6:00 AM), http://www.npr.org/2011/11/26/142758000/at-lapd-predicting-crimes-before-they-happen, *archived at* http://perma.cc/P5JL-ZVWV (discussing how police use data to predict future crimes); Joel Rubin, *Stopping Crime Before It Starts*, L.A. TIMES (Aug. 21, 2010),

Crimes are recorded.[25] Criminals are cataloged.[26] Some jurisdictions record data about every police–citizen encounter, making both the person and justification for the stop (not necessarily even an arrest) instantly available to any officer.[27] Some jurisdictions have compiled "bad guy lists" identifying suspects in a neighborhood based on computer analysis of past actions and arrests.[28] In addition, law enforcement agencies increasingly rely on predictive algorithms to forecast individual recidivism and areas of likely criminal activity.[29]

Just as law enforcement agencies now collect and electronically analyze more personal data, so do private, third-party organizations.[30] These third-party entities are a familiar part of our daily lives. "Smartphones" record

http://articles.latimes.com/2010/aug/21/local/la-me-predictcrime-20100427-1, *archived at* http://perma.cc/N223-8J9K (suggesting that predictive policing that uses sophisticated data systems is the future of law enforcement).

[25] *Cf.* Andrew Guthrie Ferguson, *Crime Mapping and the Fourth Amendment: Redrawing 'High Crime Areas,'* 63 HASTINGS L.J. 179, 225-27 (2011) [hereinafter Ferguson, *Crime Mapping*] (discussing issues with recorded crime data).

[26] *Cf. id.* at 182 n.11.

[27] For example, in New York City, every stop-and-frisk is supposed to be recorded in an official UF-250 police report. *See* Bernard E. Harcourt & Tracey L. Meares, *Randomization and the Fourth Amendment*, 78 U. CHI. L. REV. 809, 862 n.210 (2011) ("According to the NYPD's Patrol Guide, a police officer who stops and frisks an individual must complete a UF-250 if a person is (1) stopped by force; (2) stopped and frisked or searched; (3) arrested; or (4) stopped and refuses to identify oneself. . . . In situations that fall outside these four contexts, a police officer may fill out a form if he or she desires to do so." (citation omitted)).

[28] *See* Stephen D. Mastrofski & James J. Willis, *Police Organization Continuity and Change: Into the Twenty-First Century*, *in* 39 CRIME & JUSTICE 55, 88 (Michael Tonry ed., 2010) ("Police now appear to rely more heavily on certain IT-based forms of surveillance—'database policing'—where officers use computers to 'patrol' massive data files (e.g., wanted lists) looking for 'hits' on information they possess on suspects."); Bryan Llenas, *Brave New World of "Predictive Policing" Raises Specter of High-Tech Racial Profiling*, FOX NEWS LATINO (Feb. 25, 2014), http://latino.foxnews.com/latino/news/2014/02/24/brave-new-world-predictive-policing-raises-specter-high-tech-racial-profiling/, *archived at* http://perma.cc/VG5W-WV93 ("[T]he Chicago Police Department, thanks to federal funding, is now helping to drive policing into territory previously only dreamed of in science fiction: The ability to essentially predict who will be the next perpetrator or the next victim of a crime."); Robert L. Mitchell, *Predictive Policing Gets Personal*, COMPUTERWORLD, (Oct. 24, 2013, 3:50 PM), http://www.computerworld.com/article/2486424/government-it/predictive-policing-gets-personal.html, *archived at* http://perma.cc/GDW5-B8JD ("Predictive policing is at the top of a lot of people's lists.").

[29] *See* Shima Baradaran & Frank L. McIntyre, *Predicting Violence*, 90 TEX. L. REV. 497, 507 (2012) (discussing how the majority of states detain or only conditionally release defendants determined to be dangerous); Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 265-69 (2012) [hereinafter Ferguson, *Predictive Policing*] (providing an overview of predictive policing).

[30] *See infra* Part II.

where we go.[31] Credit card companies record what we buy, and banks chronicle what we spend.[32] "OnStar" systems in cars catalog where and how fast we drive.[33] Phone records reflect our contacts and communications.[34] Internet searches reveal what we read and expose our interests.[35] Social media sites, such as Twitter and Facebook, even disclose what we think.[36] Currently, law enforcement officers may access many of these records without violating the Fourth Amendment, under the theory that there is no reasonable expectation of privacy in information knowingly revealed to third parties.[37] While certain statutory protections exist, most statutes include law enforcement exceptions,[38] and in any case, these private, commercial data aggregators have turned personal data into a commodity, available for purchase and analysis to anyone willing to pay.[39]

[31] *See, e.g.*, Eric Lichtblau, *Police Are Using Phone Tracking as Routine Tool*, N.Y. TIMES, Apr. 1, 2012, at 1.

[32] *Cf., e.g.*, Charles Duhigg, *Psst, You in Aisle 5*, N.Y. TIMES, Feb. 16, 2012 (Magazine), at 30 (discussing retail analytics).

[33] *See, e.g.*, Ned Potter, *Privacy Battles: OnStar Says GM May Record Car's Use, Even If You Cancel Service*, ABC NEWS (Sept. 26, 2011), http://abcnews.go.com/Technology/onstar-gm-privacy-terms-company-record-car-information/story?id=14581571, *archived at* http://perma.cc/VGN2-MJMZ.

[34] Phone companies record whom we call and even where we are located when we make those calls. *See, e.g.*, Noam Cohen, *It's Tracking Your Every Move, and You May Not Even Know*, N.Y. TIMES, Mar. 26, 2011, at A1.

[35] *See, e.g.*, Chloe Albanesius, *Facebook: Tracking Your Web Activity Even After You Log Out?*, PC MAG. (Sept. 26, 2011, 11:59 AM), http://www.pcmag.com/article2/0,2819,2393564,00.asp, *archived at* http://perma.cc/NWP2-DRXN; Robert Epstein, *Google's Gotcha*, U.S. NEWS & WORLD REP. (May 10, 2013, 12:15 PM), http://www.usnews.com/opinion/articles/2013/05/10/15-ways-google-monitors-you, *archived at* http://perma.cc/94V8-AUSX.

[36] *See generally* Noah Shachtman, *Exclusive: U.S. Spies Buy Stake in Firm That Monitors Blogs, Tweets*, WIRED (Oct. 19, 2009, 12:03 PM), http://www.wired.com/2009/10/exclusive-us-spies-buy-stake-in-twitter-blog-monitoring-firm/, *archived at* http://perma.cc/BZC6-SWAS (highlighting the intelligence value of social media posts).

[37] *See, e.g.*, Stephen E. Henderson, *Beyond the (Current) Fourth Amendment: Protecting Third-Party Information, Third Parties, and the Rest of Us Too*, 34 PEPP. L. REV. 975, 982-83 (2007) [hereinafter Henderson, *Beyond the (Current) Fourth Amendment*] (exploring the ways in which the third-party doctrine shortchanges privacy interests); Stephen E. Henderson, *Learning from All Fifty States: How to Apply the Fourth Amendment and Its State Analogs to Protect Third Party Information from Unreasonable Search*, 55 CATH. U. L. REV. 373, 376-79 (2006) [hereinafter Henderson, *Fifty States*] (describing the third-party doctrine); Orin S. Kerr, *The Case for the Third-Party Doctrine*, 107 MICH. L. REV. 561, 563 (2009) (same).

[38] 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 & n.2 (2013) ("The United States Code currently contains over twenty separate statutes that restrict both the acquisition and release of covered information. . . . Yet across this remarkable diversity, there is one feature that all these statutes share in common: each contains a provision exempting law enforcement from its general terms.").

[39] *See infra* Part II.

The rise of "big data" means that this information is potentially available for use by law enforcement. In the same way that a drug store can predict that you will need a coupon this month because you bought a similar product last month,[40] the police will be able to anticipate that you will be selling drugs this week because you purchased an unusual number of mini–plastic bags last week.[41] Neither prediction is necessarily accurate, but both are based on individualized and particularized data that makes the prediction more likely.

Part III analyzes the intersection of big data and the current Fourth Amendment framework. The wrinkle of big data is that now officers are no longer dealing with "strangers." Even people unknown to officers can be identified and, with a few quick searches, revealed as a person with recognizable characteristics or about whom certain predictions can be made.[42] If officers view those individualized and particularized identifying characteristics—such as prior convictions, gang associations, and GPS coordinates near the scene of the crime—as suspicious, then otherwise innocent actions might create a predictive composite that satisfies the reasonable suspicion standard. In essence, reasonable suspicion will focus more on an individual's predictive likelihood of involvement in criminal activity than on an individual's actions.

Part III then looks at Fourth Amendment reasonable suspicion through three different lenses: (1) situations involving officers observing an ongoing crime, (2) situations involving officers investigating a past crime, and (3) situations involving officers predicting a future crime. Big data affects the analysis in each application, distorting the reasonable suspicion standard. Knowing who the suspect is and having more information (even innocent information) will allow officers to meet the reasonable suspicion threshold more easily because the information will be sufficiently individualized and particularized.

[40] *See* Duhigg, *supra* note 32, at 30; Rebecca Greenfield, *Facebook Now Knows What You're Buying at Drug Stores*, WIRE (Sept. 24, 2012, 11:49 AM), http://www.thewire.com/technology/2012/09/facebook-tracking-you-drug-store-now-too/57183/, *archived at* http://perma.cc/N5XH-QBA4; William F. Pewen, *Protecting Our Civil Rights in the Era of Digital Health*, ATLANTIC (Aug. 2, 2012, 11:09 AM), http://www.theatlantic.com/health/archive/2012/08/protecting-our-civil-rights-in-the-era-of-digital-health/260343/?single_page=true/, *archived at* http://perma.cc/8FB6-VERF.

[41] Mini–plastic bags (e.g., Ziploc bags) are used to package drugs sold on the street including cocaine, heroin, and marijuana. *See* United States v. Dingle, 114 F.3d 307, 309 (D.C. Cir. 1997) ("The government's narcotics expert testified that crack cocaine is typically packaged in small ziplock bags for street-level distribution."); United States v. Betts, 16 F.3d 748, 757 (7th Cir. 1994) (noting that pagers and Ziploc baggies are "hallmark paraphernalia" of drug distribution).

[42] *See infra* Part III.

Part IV assesses this new technological reality. Can the current reasonable suspicion doctrine adapt? Should it? What are the possible benefits or dangers of big data reasonable suspicion? Using big data may help reduce the negative consequences of traditional policing techniques, but at the same time may create a whole new set of concerns. This section evaluates the tradeoffs of big data as applied to the Fourth Amendment.

Part V offers a few solutions to the problem presented by the big data distortions of Fourth Amendment doctrine. This Article suggests that the nature of big data itself might provide a means of strengthening the reasonable suspicion standard. If big data resources are used to tip the scales of reasonable suspicion in favor of law enforcement, then courts should require a higher level of detail and correlation using the insights and capabilities of big data. This requirement would involve precise statistical analysis, geospatial analysis, temporal analysis, and link analysis of the data. Big data can provide information about a person on a generalized or granular scale, and the latter should be required. The power of big data allows investigators to go deep into the data and make sure that the information is as tightly correlated as possible. In this way, a big data–infused reasonable suspicion standard will do what the reasonable suspicion requirement was always supposed to do—distinguish the criminal from the noncriminal in a manner that balances the need for effective law enforcement with a measure of personal liberty.

I. REASONABLE SUSPICION: A SMALL DATA DOCTRINE

The Fourth Amendment serves as a constitutional barrier, protecting individuals from unreasonable police intrusion.[43] On the street, the police may not constitutionally stop, seize, or search individuals without the requisite legal justification.[44] To seize a person temporarily, a police officer

[43] U.S. CONST. amend. IV; Dunaway v. New York, 442 U.S. 200, 213 (1979) ("Hostility to seizures based on mere suspicion was a prime motivation for the adoption of the Fourth Amendment, and decisions immediately after its adoption affirmed that 'common rumor or report, suspicion, or even "strong reason to suspect" was not adequate to support a warrant for arrest.'" (citing Henry v. United States, 361 U.S. 98, 101 (1959))); Almeida-Sanchez v. United States, 413 U.S. 266, 273 (1973) ("The needs of law enforcement stand in constant tension with the Constitution's protections of the individual against certain exercises of official power. It is precisely the predictability of these pressures that counsels a resolute loyalty to constitutional safeguards.").

[44] *See* Brown v. Texas, 443 U.S. 47, 51 (1979) ("A central concern in balancing these competing considerations in a variety of settings has been to assure that an individual's reasonable expectation of privacy is not subject to arbitrary invasions solely at the unfettered discretion of officers in the field.").

must have "reasonable suspicion" that the individual is committing, is about to commit, or has committed a crime.[45]

The "reasonable suspicion" standard first arose in *Terry v. Ohio*, when the Supreme Court created a new threshold for Fourth Amendment suspicion, lower than probable cause, to justify a brief detention.[46] In *Terry*, Detective Martin McFadden observed three unknown men walking back and forth in front of a downtown store.[47] McFadden, an experienced police officer, while not knowing the men involved, believed their actions were consistent with the actions of individuals seeking to rob a store.[48] Based on this suspicion, McFadden stopped the individuals.[49] In the process of frisking them, McFadden recovered unlawful firearms.[50] Possession of these firearms served as the basis for the arrest, conviction, and later appeal of the constitutionality of the initial stop-and-frisk.[51] In finding the stop permissible, the Court established a new Fourth Amendment standard for investigatory stops, requiring that police "be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant th[e] intrusion."[52]

Terry mirrors a common factual situation that recurs millions of times a year across the country. Officers on the street observe a particular suspect or group of suspects. Because police officers do not know all of the suspects in a jurisdiction personally, the officers must focus on the suspect's actions and on the inferences that can be drawn from those actions. The amount of information an officer knows about the suspect is necessarily limited. Like Detective McFadden in *Terry*, officers are usually limited to evaluating observed actions through intuition. This practice epitomizes small data

[45] *See* Terry v. Ohio, 392 U.S. 1, 21-22 (1968) (articulating the standard and explaining the rationale behind it); *see also* United States v. Hensley, 469 U.S. 221, 227 (1985) (highlighting cases where the Court upheld investigatory stops where police had less than probable cause but nonetheless had reasonable suspicion that a crime was in progress, would be committed, or had been committed).

[46] 392 U.S. at 21-22.

[47] *Id.* at 5-6.

[48] Of course, McFadden admitted at the trial level that he had no significant experience investigating robbery suspects, and the overlay of racial considerations in a segregated area of downtown Cleveland in the late 1960s likely contributed to his suspicion. *See* Thomas B. McAffee, *Setting Us Up for Disaster: The Supreme Court's Decision in* Terry v. Ohio, 12 NEV. L.J. 609, 611 n.13 (2012) (discussing the racial context of Cleveland at the time); *see also* Terry v. Ohio *30 Years Later: A Symposium on the Fourth Amendment, Law Enforcement and Police–Citizen Encounters*, 72 ST. JOHN'S L. REV. 721 app. B, at 1477 (John Q. Barrett ed., 1998) [hereinafter Terry v. Ohio *30 Years Later*] (reporting Detective McFadden's testimony).

[49] *See Terry*, 392 U.S. at 7-8.

[50] *Id.*

[51] *Id.*

[52] *Id.* at 21.

policing—suspicion generated by information discrete in amount, fixed in time, and isolated in context. Thus, the predictive judgments made about the suspect are similarly limited and disconnected from other data sources.

This Part discusses how reasonable suspicion has developed as a small data doctrine. The language the Supreme Court used to define reasonable suspicion, the standard's application in a variety of contexts, and the law's assumption of unknown suspects and direct observations, all speak to the doctrine's utility in certain situations. In general, the archetypical reasonable suspicion setting involves police officers reacting to a quickly unfolding criminal situation, with unknown suspects and without the time or resources to find more information. The Fourth Amendment calculus, though, changes when officers have access to personal data about the suspects. Specifically, law enforcement officers can more easily satisfy the reasonable suspicion standard when a third party provides some minimal information about an otherwise unknown suspect.

Already, the reasonable suspicion standard provides little protection in situations involving suspects previously known to police. This reality illustrates one of the many shortcomings of a small data doctrine. Specific and particularized data about *a suspect*, even if not specific and particularized about *a crime*, tends (in practice) to reduce the protection of the reasonable suspicion standard. Because the police can obtain information about a suspect more easily in a world of big data, this doctrinal weakness points to a problem in the protective scope of current Fourth Amendment law.

A. *The Reasonable Suspicion Standard*

Despite the common application of the reasonable suspicion language in tens of thousands of federal and state court cases, the contours of the standard remain ill-defined.[53] Cases from *Terry* to the present day emphasize that suspicion must be based on "specific and articulable facts."[54] Those facts must be "objective."[55] Suspicion must be particularized.[56] It must

[53] *See* William J. Stuntz, *Local Policing After the Terror*, 111 YALE L.J. 2137, 2175 (2002) ("The central problem with regulating the manner of street stops is definition: No one knows how to craft a legal formula that will tell officers how to behave in advance.").

[54] *See Terry*, 392 U.S. at 21 (holding that an officer must be able to identify "specific and articulable facts[,] which . . . together with rational inferences from those facts," establish the reasonable suspicion of criminal activity required to justify a seizure).

[55] Brown v. Texas, 443 U.S. 47, 51 (1979) (noting that police officers must "have a reasonable suspicion, based on objective facts, that the individual is involved in criminal activity" to make a investigatory stop); *see also id.* ("To this end, the Fourth Amendment requires that a seizure must be based on specific, objective facts indicating that society's legitimate interests require the seizure

relate to criminal activity, not just to the criminal.[57] Officers must have reasonable suspicion before the stop occurs; retroactive justification is not sufficient.[58] The suspicion must relate to current criminal activity,[59] with some latitude for post-crime investigative actions[60] and pre-crime intervention.[61]

Courts evaluate these suspicious facts under the "totality of circumstances" test, which means that all relevant factors should be considered.[62] The content and the quality of the information are both relevant considerations,[63] but courts have not settled on a required quantum of proof.[64]

of the particular individual, or that the seizure must be carried out pursuant to a plan embodying explicit, neutral limitations on the conduct of individual officers.").

[56] *See* United States v. Arvizu, 534 U.S. 266, 270-72 (2002) (noting that reasonable suspicion must be based on a "'[p]articularized and objective basis' for suspecting legal wrongdoing" (quoting United States v. Cortez, 449 U.S. 411, 417 (1981))).

[57] Florida v. J.L., 529 U.S. 266, 272 (2000) ("The reasonable suspicion here at issue requires that a tip be reliable in its assertion of illegality, not just in its tendency to identify a determinate person."); United States v. Cortez, 449 U.S. 411, 417 (1981) ("An investigatory stop must be justified by some objective manifestation that the person stopped is, or is about to be, engaged in criminal activity.").

[58] *J.L.*, 529 U.S. at 271 ("The reasonableness of official suspicion must be measured by what the officers knew before they conducted their search.").

[59] *See* United States v. Sokolow, 490 U.S. 1, 12 (1989) (Marshall, J., dissenting) ("It is not enough to suspect that an individual has committed crimes in the past, harbors unconsummated criminal designs, or has the propensity to commit crimes. On the contrary, before detaining an individual, law enforcement officers must reasonably suspect that he is engaged in, or poised to commit, a criminal act *at that moment*.").

[60] *See* United States v. Hensley, 469 U.S. 221, 227-29 (1985) (upholding a stop based only on a "wanted flyer" from another police department that named the suspect).

[61] *See generally* Jennifer C. Daskal, *Pre-Crime Restraints: The Explosion of Targeted, Noncustodial Prevention*, 99 CORNELL L. REV. 327, 335-357 (2014) (discussing restraints that target yet-uncommitted crime, like the No-Fly List and Megan's Laws).

[62] *See* Alabama v. White, 496 U.S. 325, 330 (1990) (explaining that reasonable suspicion depends on the totality of the circumstances); *Sokolow*, 490 U.S. at 7-8 ("The concept of reasonable suspicion, like probable cause, is not 'readily, or even usefully, reduced to a neat set of legal rules.' . . . In evaluating the validity of a stop such as this, we must consider 'the totality of the circumstances—the whole picture.'" (quoting Illinois v. Gates, 462 U.S. 213, 232 (1983); *Cortez*, 449 U.S. at 417)).

[63] *White*, 496 U.S. at 330 ("Reasonable suspicion, like probable cause, is dependent upon both the content of information possessed by police and its degree of reliability. Both factors—quantity and quality—are considered in the 'totality of the circumstances—the whole picture,' that must be taken into account when evaluating whether there is reasonable suspicion." (citation omitted) (quoting *Cortez*, 449 U.S. at 417)).

[64] *See, e.g.*, Illinois v. Wardlow, 528 U.S. 119, 123 (2000) ("While 'reasonable suspicion' is a less demanding standard than probable cause and requires a showing considerably less than preponderance of the evidence, the Fourth Amendment requires at least a minimal level of objective justification for making the stop.").

Innocent factors, characteristics about an area, and specialized law enforcement training are all factors that shape the totality of the circumstances.[65]

The result has been a standard which retains the virtue of flexibility and the vice of malleability. As the Supreme Court has explained, reasonable suspicion involves "commonsense, nontechnical conceptions that deal with 'the factual and practical considerations of everyday life on which reasonable and prudent men, not legal technicians, act.'"[66] Scholars have been less charitable, critiquing the standard as being at best meaningless and at worst discriminatory.[67] At a minimum, the reasonable suspicion standard requires police to articulate why an individual was stopped, which may reduce overly arbitrary or animus-based stops.

The standard, while applicable in many situations, makes the most sense for unknown suspect cases. The requirement of specific, particularized, objective facts seeks to distinguish by their observable actions those individuals who have done nothing wrong from those who have done something wrong. Detective McFadden stopped Mr. Terry not because he recognized Mr. Terry as a known robber or because there was a report of a robbery, but because of the observed actions that drew his suspicions at that particular time. Personal factors may be relevant, but usually relate to the suspect's observable actions.[68]

[65] *See, e.g.*, United States v. Mendenhall, 446 U.S. 544, 563 (1980) ("Among the circumstances that can give rise to reasonable suspicion are the agent's knowledge of the methods used in recent criminal activity and the characteristics of persons engaged in such illegal practices.").

[66] Ornelas v. United States, 517 U.S. 690, 695 (1996) (quoting *Gates*, 462 U.S. at 231).

[67] *See, e.g.*, David A. Harris, *Frisking Every Suspect: The Withering of* Terry, 28 U.C. DAVIS L. REV. 1, 5 n.18 (1994) (noting that for certain minority groups, reasonable suspicion seems to include even benign conduct); David A. Harris, *Particularized Suspicion, Categorical Judgments: Supreme Court Rhetoric Versus Lower Court Reality Under* Terry v. Ohio, 72 ST. JOHN'S L. REV. 975, 1022 (1998) (arguing that the reasonable suspicion standard has led to targeting of minorities for stops almost at will); Lewis R. Katz, Terry v. Ohio *at Thirty-Five: A Revisionist's View*, 74 MISS. L.J. 423, 493 (2004) ("[I]n . . . the inner city, the possibility of criminal activity is so substantial as to make everyone in the area subject to police inquiry."); Christopher Slobogin, *Let's Not Bury* Terry*: A Call for Rejuvenation of the Proportionality Principle,* 72 ST. JOHN'S L. REV. 1053, 1081 (1998) (calling reasonable suspicion jurisprudence "a mess").

[68] For example, the "high crime area" nature of a neighborhood has been accepted as a contextual factor that may affect reasonable suspicion. *See, e.g.*, Andrew Guthrie Ferguson & Damien Bernache, *The "High Crime Area" Question: Requiring Verifiable and Quantifiable Evidence for Fourth Amendment Reasonable Suspicion Analysis*, 57 AM. U. L. REV. 1587, 1609-18 (2008) ("[W]hat is termed a 'high-crime area' can differ from case to case, and jurisdiction to jurisdiction."); Margaret Raymond, *Down on the Corner, Out in the Street: Considering the Character of the Neighborhood in Evaluating Reasonable Suspicion*, 60 OHIO ST. L.J. 99, 120-22 (1999) ("[S]tanding on a street corner may create reasonable suspicion in Louisiana, but not in Pennsylvania . . . .").

B. *Reasonable Suspicion in Application*

This Section traces how an officer's knowledge about a suspect influences the reasonable suspicion analysis. As discussed below, the more data known or discovered about a particular suspect, the easier it is to justify a stop based on reasonable suspicion. This result is not necessarily negative or surprising, as more information connecting a suspect to a crime increases the likelihood that the suspect was involved. It does, however, show how the aggregation of information—even innocent information—can shift the balance within a fluid legal standard.

1. Unknown Suspect Cases

Perhaps not surprisingly, many Fourth Amendment reasonable suspicion cases have followed the *Terry* small data model.[69] Such cases involve officers observing unknown suspects[70] engaged in what officers believe to be suspicious activity. These encounters regularly occur on the street[71] and as part of traffic stops that lead to the seizure of cars' occupants.[72] In both types of encounters, police stop the individual based on speculations about the criminal nature of the actions involved—independent of the person—because the actor's identity is unknown. In determining reasonable suspicion,

[69] *See, e.g.*, United States v. Arvizu, 534 U.S. 266, 268-72 (2002) (involving an unknown driver stopped on suspicion of drug trafficking); United States v. Sharpe, 470 U.S. 675, 677-78 (1985) (same); Florida v. Rodriguez, 469 U.S. 1, 2-4 (1984) (per curiam) (involving unknown suspects stopped because of suspicious and furtive movements in an airport); Florida v. Royer, 460 U.S. 491, 493 (1983) (involving an unknown traveler stopped in an airport because his actions fit a "drug courier profile"); United States v. Brignoni-Ponce, 422 U.S. 873, 875 (1975) (involving unknown suspects believed to be in the country unlawfully); Adams v. Williams, 407 U.S. 143, 144-45 (1972) (involving an unknown suspect stopped based on an informant's tip that the suspect was armed).

[70] Specifically, suspects personally unknown to the officer.

[71] *See* Brown v. Texas, 443 U.S. 47, 49 (1979) (involving an unknown suspect stopped in an alley in a neighborhood known for drug trafficking); *see also* Illinois v. Wardlow, 528 U.S. 119, 121 (2000) (involving a stop of an unknown suspect who fled on foot after seeing police in a Chicago neighborhood known for drug trafficking); California v. Hodari D., 499 U.S. 621, 622-23 (1991) (involving a chase of unknown juvenile suspects who fled on foot after seeing police).

[72] *See, e.g.*, Bailey v. United States, 133 S. Ct. 1031, 1036 (2013) (involving the stop of an automobile occupied by unknown suspects who had just left an apartment for which the police had obtained a search warrant); Arizona v. Johnson, 555 U.S. 323, 327 (2009) (involving a traffic stop for suspended vehicle registration); Brendlin v. California, 551 U.S. 249, 252 (2007) (involving a traffic stop stemming from vehicle registration concerns); Illinois v. Caballes, 543 U.S. 405, 406 (2005) (involving a traffic stop for speeding); Knowles v. Iowa, 525 U.S. 113, 114 (1998) (same); Ohio v. Robinette, 519 U.S. 33, 35 (1996) (same); New York v. Class, 475 U.S. 106, 107-08 (1986) (involving a traffic stop of an unknown suspect for speeding and driving with a cracked windshield).

a court will evaluate whether the officer's observations were objectively reasonable to warrant the stop.

Sibron v. New York is a useful example of the application of the reasonable suspicion standard with respect to unknown suspects.[73] In *Sibron*, a case decided the same day as *Terry v. Ohio*, an officer observed an unknown suspect in a series of meetings with known narcotics addicts for approximately eight hours.[74] The police officer did not know Sibron personally and could not overhear any of the conversations.[75] Yet, after observing Sibron communicate with nine to eleven known addicts over the course of the day, the officer concluded that he had observed a series of drug transactions. The officer then approached Sibron and searched his pockets, recovering heroin.[76] The Supreme Court held that Sibron's activity did not create reasonable suspicion.[77] The officer did not see any drug transactions, did not know the subject of the conversations, and accordingly did not have the requisite suspicion to justify the search under the Fourth Amendment.[78] While

[73] 392 U.S. 40 (1968).

[74] *Id.* at 45 ("Officer Martin testified that while he was patrolling his beat in uniform on March 9, 1965, he observed Sibron 'continually from the hours of 4:00 P.M. to 12:00, midnight . . . in the vicinity of 742 Broadway.' He stated that during this period of time he saw Sibron in conversation with six or eight persons whom he (Patrolman Martin) knew from past experience to be narcotics addicts. The officer testified that he did not overhear any of these conversations, and that he did not see anything pass between Sibron and any of the others. Late in the evening Sibron entered a restaurant. Patrolman Martin saw Sibron speak with three more known addicts inside the restaurant. Once again, nothing was overheard and nothing was seen to pass between Sibron and the addicts. Sibron sat down and ordered pie and coffee, and, as he was eating, Patrolman Martin approached him and told him to come outside. Once outside, the officer said to Sibron, 'You know what I am after.' According to the officer, Sibron 'mumbled something and reached into his pocket.' Simultaneously, Patrolman Martin thrust his hand into the same pocket, discovering several glassine envelopes, which, it turned out, contained heroin." (alteration in original)).

[75] *Id.*

[76] *Id.*

[77] *Id.* at 64. Because *Sibron* was decided on the same day as *Terry*, there is some debate about whether the Court found the search unconstitutional because the officer lacked probable cause to search or because the police lacked reasonable suspicion to stop-and-frisk. The majority opinion echoed the reasonable suspicion language in *Terry* without specifically using the term. "The inference that persons who talk to narcotics addicts are engaged in the criminal traffic in narcotics is simply not the sort of reasonable inference required to support an intrusion by the police upon an individual's personal security." *Id.* at 62. Justice Harlan more explicitly referenced the *Terry* standard in his concurrence. "The forcible encounter between Officer Martin and Sibron did not meet the *Terry* reasonableness standard. In the first place, although association with known criminals may, I think, properly be a factor contributing to the suspiciousness of circumstances, it does not, entirely by itself, create suspicion adequate to support a stop." *Id.* at 73 (Harlan, J., concurring).

[78] *Id.* at 62 ("The officer was not acquainted with Sibron and had no information concerning him. He merely saw Sibron talking to a number of known narcotics addicts over a period of eight hours. It must be emphasized that Patrolman Martin was completely ignorant regarding the

Sibron might be decided differently today, the Court, it should be noted here, determined that the officer lacked reasonable suspicion because the officer had no information about Sibron that gave rise to the inference of criminal activity. Merely associating with addicts was not a crime.

In both *Terry* and *Sibron*, officers did not know the suspect but instead inferred from the unknown suspect's actions that criminal activity was afoot.[79] In both cases, officers based their predictive judgments on limited data points, which were disconnected from larger information sources about the suspect. While the Supreme Court came to different conclusions in *Terry* and *Sibron* about whether the police had reasonable suspicion, it reached both outcomes based solely on the facts observed by the officers. This reality has been repeated in hundreds of cases since then.

2. Some Information on the Suspect

A slight wrinkle to the classic unknown suspect case involves situations in which some minimal information is provided about an otherwise unknown suspect. Informant tips, police tips, or further police investigation can alter the reasonable suspicion analysis.[80] That is, the likelihood that a court will find reasonable suspicion increases proportionally to the amount of personal data the police officer has about the suspect.

In *Florida v. J.L.*, an anonymous caller "reported to the Miami–Dade Police that a young black male standing at a particular bus stop and wearing a plaid shirt was carrying a gun."[81] Police responded and searched a young black male, J.L., wearing a plaid shirt. Police found a gun on J.L. and arrested him.[82] In determining whether this anonymous tip was sufficient to justify the stop, the Court held that such a bare tip, without either identifying information about the suspect or predictive details corroborated by observation, was insufficient for reasonable suspicion.[83] There was no identifying or

content of these conversations, and that he saw nothing pass between Sibron and the addicts. So far as he knew, they might indeed 'have been talking about the World Series.'").

[79] *See* United States v. Sokolow, 490 U.S. 1, 7 (1989) ("[P]olice can stop and briefly detain a person for investigative purposes if the officer has a reasonable suspicion supported by articulable facts that criminal activity 'may be afoot,' even if the officer lacks probable cause.").

[80] *See, e.g.*, Hiibel v. Sixth Judicial Dist. Court of Nev., 542 U.S. 177, 180-81 (2004) (involving a telephone tip that provided reasonable suspicion to investigate a domestic violence case).

[81] 529 U.S. 266, 268 (2000).

[82] *Id.* at 268-69.

[83] *Id.* at 270 ("Unlike a tip from a known informant whose reputation can be assessed and who can be held responsible if her allegations turn out to be fabricated, . . . 'an anonymous tip alone seldom demonstrates the informant's basis of knowledge or veracity.'" (quoting Alabama v. White, 496 U.S. 325, 329 (1990))).

predictive information involved. The only data point was an anonymous accusation without context or verifiability.

In *Alabama v. White*, by contrast, the Supreme Court found that a tip including the suspect's name, location, and predicted route of travel was sufficient to establish reasonable suspicion.[84] In *White*, the anonymous tipster claimed that the suspect "would be leaving 235-C Lynwood Terrace Apartments at a particular time in a brown Plymouth station wagon with the right taillight lens broken, that she would be going to Dobey's Motel, and that she would be in possession of about an ounce of cocaine inside a brown attaché case."[85] The police followed the suspect as she left the motel in a Plymouth station wagon, stopped her, and requested to search her car.[86] She consented, and police recovered marijuana.[87] The Court held that in this situation, the stop was justified by reasonable suspicion.[88]

The differences between *White* and *J.L.* are slight but revealing. A more descriptive account of the suspect and corroborated predictive detail changed an insufficient anonymous tip into constitutionally sufficient reasonable suspicion.[89] Note, however, that many of the same concerns that caused the Court to find no reasonable suspicion in *J.L.* were still present in *White*. An anonymous tip revealing a single and obvious pattern of movement does not provide much proof of "insider" knowledge. Most individuals drive a particular type of car and follow a predictable routine in parts of daily life (e.g., driving to work, to daycare, to the gym, to the local coffee shop). Yet, the Court still found the additional individualized information about the suspect sufficient for reasonable suspicion.[90]

The Supreme Court's first transition to a "medium data" case occurred in *Ornelas v. United States*.[91] In *Ornelas*, officers developed reasonable suspicion by proactively searching for data to support their hunch.[92] The

[84] 496 U.S. at 331-32.

[85] *Id.* at 327.

[86] *Id.*

[87] *Id.*

[88] *Id.* at 332.

[89] *Compare id.* ("Although it is a close case, we conclude that under the totality of the circumstances the anonymous tip, as corroborated, exhibited sufficient indicia of reliability to justify the investigatory stop of respondent's car."), *and* Florida v. J.L., 529 U.S. 266, 270 (2000) ("Only after police observation showed that the informant had accurately predicted the woman's movements, we explained, did it become reasonable to think the tipster had inside knowledge about the suspect and therefore to credit his assertion about the cocaine."), *with id.* at 271 ("The anonymous call concerning J.L. provided no predictive information and therefore left the police without means to test the informant's knowledge or credibility.").

[90] *See White*, 496 U.S. at 332.

[91] 517 U.S. 690 (1996).

[92] *Id.* at 692.

case involved a Milwaukee detective who observed a suspicious car parked at a local motel.[93] The car was purportedly "suspicious" because it was a make and model frequently used by drug dealers—a 1981 two-door Oldsmobile.[94] Not having any information about the owner of the Oldsmobile, the detective radioed his dispatcher and found the car was registered under Ornelas's name.[95] A further inquiry with the local office of the Drug Enforcement Administration revealed that Ornelas's name appeared in a federal database of known and suspected drug traffickers (the Narcotics and Dangerous Drugs Information System (NADDIS)).[96] With this additional information, which would otherwise have been unknown, the detective stopped Ornelas (and another man) when they exited the motel and entered the car. The underlying constitutional issue in the case was whether the information connecting the car, the name, and the criminal database was sufficient to create reasonable suspicion.[97] While the Supreme Court deferred answering the Fourth Amendment question, focusing instead on the appropriate standard of appellate review, the trial court both initially and on remand found that the information together was sufficient cause for reasonable suspicion.[98] Note, though, that Ornelas's actions were not suspicious at all. He parked overnight at a motel and then exited the motel and got into his car. What created the suspicion was independent data about Ornelas himself. The detective, by searching for more information about

[93] *Id.* at 691-92.

[94] *Id.* at 692.

[95] *Id.*

[96] *Id.*

[97] *Id.* at 694.

[98] *See* United States v. Ornelas, No. 94-3349, 1996 WL 508569, at *1 (7th Cir. Sept. 4, 1996). The Seventh Circuit, however, raised concerns with the accuracy and use of the information. *See* United States v. Ornelas-Ledesma, 16 F.3d 714, 716-17 (7th Cir. 1994) ("Clearly, were it not for the NADDIS hits, the officers would not have had grounds for *reasonable* suspicion that the defendants were drug traffickers. Not only is every circumstance on which the officers relied other than the hits innocent taken by itself—many Americans (approximately one in eight) are Californians, many Californians are Hispanic, many Americans drive two-door General Motors cars, many people check into motels very late at night (or early in the morning), many travel in pairs rather than alone, and many do not make advance reservations—but the confluence of these circumstances is pretty innocuous as well, especially since many of the circumstances are correlated rather than independent."); *see also id.* at 717 ("Maybe NADDIS is no better than a vast compendium of rumors, errors, and libels: garbage in, garbage out. That seems unlikely. It would not be heavily used by drug enforcement authorities if it were merely a random sample of the American population. Which is not to say, however, that it is *highly* reliable; concern that it may not be is heightened by the (scanty) secondary literature, which depicts NADDIS as an unselective, unweeded repository of unsubstantiated allegations, often dated.").

Ornelas, discovered personal data—particularized and individualized facts—to support his suspicion.[99]

This development of reasonable suspicion resulting more from aggregated information about a suspect, and less from the actions of the suspect, occurs with some regularity in the case law.[100] In essence, courts reason that the "tip" or database hit provides information that shifts the balance toward reasonable suspicion. This is true even if the observable, innocent actions on the street remain the same.

3. Known Suspects

The third situation involves "known suspects" stopped by police because of their identity and not necessarily because of any observed activities. The Supreme Court has not directly ruled on the issue, but *United States v. Hensley* provides an interesting example of a stop based solely on identity.[101]

Hensley involved a stop based on a "wanted flyer" for a suspect in an armed robbery.[102] The Police Department in St. Bernard, Ohio, had issued a flyer identifying Mr. Hensley as a suspect and sent it to surrounding jurisdictions.[103] The flyer did not indicate that the police had a warrant for Mr. Hensley's arrest.[104] The Covington Police Department, located in neighboring Kentucky, received the flyer, and its officers were on the lookout for Mr. Hensley.[105] Based on the wanted flyer, Covington Police

[99] *See Ornelas*, 517 U.S. at 692.

[100] *See, e.g.*, United States v. Sokolow, 490 U.S. 1, 4 (1989) (recounting an incident where an airline ticket agent informed police of a passenger's suspicious behavior, police determined that the name the suspect had given the ticket agent did not match phone records but that the suspect's voice appeared on the answering machine connected to the phone number the suspect provided); United States v. Montoya de Hernandez, 473 U.S. 531, 533 (1985) (describing details giving rise to reasonable suspicion by customs officers, including the suspect's eight prior trips to Miami or Los Angeles, possession of $5000 in cash, and lack of specific plans for her stay in the United States); *id.* at 542 ("The facts, and their rational inferences, known to customs inspectors in this case clearly supported a reasonable suspicion that respondent was an alimentary canal smuggler. We need not belabor the facts, including respondent's implausible story, that supported this suspicion. The trained customs inspectors had encountered many alimentary canal smugglers and certainly had more than an 'inchoate and unparticularized suspicion or "hunch,"' that respondent was smuggling narcotics in her alimentary canal." (citations omitted) (quoting Terry v. Ohio, 392 U.S. 1, 27 (1968)).

[101] 469 U.S. 221, 223 (1985) (addressing whether a stop based only on a "wanted flyer" from another jurisdiction runs afoul of the Fourth Amendment); *see also* Minnesota v. Dickerson, 508 U.S. 366, 368 (1993) (discussing the stop of an unknown suspect based on detective's knowledge that the building was a notorious "crack house").

[102] *Hensley*, 469 U.S. at 223.

[103] *Id.* at 225.

[104] *See id.* at 225.

[105] *Id.* at 223.

officers stopped Mr. Hensley and eventually recovered a handgun.[106] The Supreme Court had to decide whether this wanted poster—identifying Mr. Hensley specifically—created reasonable suspicion to stop Hensley.

In its discussion, the Court noted that "if police have a reasonable suspicion, grounded in specific and articulable facts, that a person they encounter was involved in or is wanted in connection with a completed felony, then a *Terry* stop may be made to investigate that suspicion."[107] The principle applies even if another jurisdiction's police officers generated that reasonable suspicion. Thus, the Covington police permissibly relied on the St. Bernard's Police Department's determination of reasonable suspicion.

Hensley shows that a stop can be based simply on identifying information about a suspect that is provided to police. In *Hensley*, the arresting officers did not have an arrest warrant or any predictive detail about Hensley's future actions, and they did not corroborate any of the allegations of criminal activity. The only data point for suspicion was Hensley's identity. Yet the Court nonetheless held that if information about an identified suspect rises to the level of reasonable suspicion, a stop is justified.[108] Because police collect a significant amount of information about suspects and create regular "target lists" of potential suspects, this type of stop, based merely on identity, raises serious questions.

Hensley also substantially broadened the application of the reasonable suspicion standard from preventing or apprehending ongoing criminal activity to investigating it after the fact. After *Hensley*, reasonable suspicion was no longer limited to ongoing criminal action, but could be used to justify stops to investigate completed crimes.

Other courts have been even more explicit that prior knowledge of the suspect can factor into reasonable suspicion. The Seventh Circuit stated that "[k]nowledge of gang association and recent relevant criminal conduct, while of doubtful *evidentiary* value in view of the strictures against proving guilt by

[106] *Id.* at 223-25.

[107] *Id.* at 229.

[108] *Id.* This recognition is also implicit (albeit grounded in a different rationale) in the Court's decisions involving the Fourth Amendment rights of probationers and individuals on parole who were stopped because their probation or parole officers knew they had prior criminal charges. *See, e.g.*, Samson v. California, 547 U.S. 843, 846 (2006) ("Officer Alex Rohleder of the San Bruno Police Department observed petitioner walking down a street with a woman and a child. Based on a prior contact with petitioner, Officer Rohleder was aware that petitioner was on parole and believed that he was facing an at-large warrant. Accordingly, Officer Rohleder stopped petitioner and asked him whether he had an outstanding parole warrant."); United States v. Knights, 534 U.S. 112, 115 (2001) ("Detective Hancock decided to conduct a search of Knights' apartment. Detective Hancock was aware of the search condition in Knights' probation order and thus believed that a warrant was not necessary.").

association or by a predisposition based on past criminal acts, is a permissible component of the articulable suspicion required for a *Terry* stop."[109] Courts in Massachusetts,[110] Minnesota,[111] and Hawaii,[112] among others,[113] recognize that knowledge of a defendant's criminal history can factor into the reasonable suspicion analysis. As one Massachusetts court reasoned,

> [t]he officers were also entitled to consider their personal knowledge of the defendant, including the fact that he had a pending court case involving charges of firearm possession and armed assault with intent to murder. In several cases, this court has allowed police knowledge of a person's arrest record or unspecified "criminal record" to be considered in a reasonable suspicion evaluation.[114]

[109] United States v. Feliciano, 45 F.3d 1070, 1074 (7th Cir. 1995).

[110] *See, e.g.*, Roe v. Att'y Gen., 750 N.E.2d 897, 914 (Mass. 2001) ("A person's prior criminal record is a legitimate factor to consider in determining whether there is reasonable suspicion for a stop or probable cause for a search or an arrest."); Commonwealth v. Dasilva, 849 N.E.2d 249, 253 (Mass. App. Ct. 2006) (allowing police knowledge of a suspect's criminal record to be considered in the reasonable suspicion evaluation); Commonwealth v. Calderon, 681 N.E.2d 1246, 1248 (Mass. App. Ct. 1997) (indicating that knowledge of defendant's criminal history can be factored into reasonable suspicion determination).

[111] *See, e.g.*, State v. Gilchrist, 299 N.W.2d 913, 916 (Minn. 1980) (confirming that knowledge of a suspect's potential involvement in a homicide and in a firearms-related incident can contribute to reasonable suspicion); State v. Bellikka, 490 N.W.2d 660, 663 (Minn. Ct. App. 1992) (holding that an officer's knowledge that a suspect had a history of burglary offenses strengthened the officer's reasonable suspicion that the suspect was involved in a recent burglary); State v. Munoz, 385 N.W.2d 373, 376 (Minn. Ct. App. 1986) (finding the officer's knowledge of the suspect's prior drug deals, possession of guns, and previous felony convictions corroborated a tip that the suspect was selling methamphetamine).

[112] *See* State v. Spillner, 173 P.3d 498, 507 (Haw. 2007) ("[A]lthough we have already emphasized that a person's prior history of drug arrests is insufficient to establish probable cause, awareness of past arrests may, when combined with other specific articulable facts indicating the probability of current criminal activity, factor into a determination that reasonable suspicion, sufficient to warrant a temporary investigate stop, exists." (quoting State v. Kaleohano, 56 P.3d 138, 148 (Haw. 2002))).

[113] *See, e.g.*, *In re* J.T., 678 S.E.2d 111, 114 (Ga. Ct. App. 2009) ("[B]ecause of their prior contact with J.T., the officers knew that J.T. was enrolled in and supposed to be attending school on the date and time in question. The officers therefore had a reasonable, particularized and objective basis for suspecting that J.T. was truant and were consequently justified in stopping him in order to determine why he was not attending school."); State v. Valentine, 636 A.2d 505, 510-11 (N.J. 1994) ("Moreover, a police officer's knowledge of a suspect's criminal history, especially where that history involves weapons offenses, is a relevant factor in judging the reasonableness of a *Terry* frisk. Although an officer's knowledge of a suspect's criminal history alone is not sufficient to justify the initial stop of a suspect or to justify a frisk of a suspect once stopped, an officer's knowledge of a suspect's prior criminal activity in combination with other factors may lead to a reasonable suspicion that the suspect is armed and dangerous.").

[114] *Dasilva*, 849 N.E.2d at 253.

Of course, courts draw a line between using prior knowledge about a suspect to justify a stop and using prior knowledge as one factor in the totality of the circumstances.[115] Prior knowledge of past criminal activity alone is not enough to stop an individual.[116] As one court stated,

> knowledge of a person's prior criminal involvement (to say nothing of a mere arrest) is alone insufficient to give rise to the requisite reasonable suspicion. . . .
>
> If the law were otherwise, any person with any sort of criminal record—or even worse, a person with arrests but no convictions—could be subjected to a *Terry*-type investigative stop by a law enforcement officer at any time without the need for any other justification at all. Any such rule would clearly run counter to the requirement of a *reasonable* suspicion, and of the need that such stops be justified in light of a balancing of the competing interests at stake.[117]

Thus, knowledge about the suspect cannot alone justify a stop; the officer's knowledge must be tied to a suspected criminal activity, past or present. Data regarding a suspect's criminal history, however, can influence the officer and be included in the totality of the circumstances analysis for reasonable suspicion.

C. *Concluding Thoughts*

Fourth Amendment case law suggests that personal information about a suspect influences the reasonable suspicion analysis—even if the suspect's actions remain the same. When data about the suspect corroborates suspicion from observation, the information helps the officer justify his or her suspicion. In simple terms, personal data provide the individualized, objective facts that officers need to articulate their suspicion.

While it makes intuitive sense that information about a suspect in connection with a crime can help provide reasons for the officer's suspicion,

[115] *See, e.g.*, United States v. Laughrin, 438 F.3d 1245, 1247 (10th Cir. 2006) (emphasizing that prior criminal involvement alone is insufficient to establish reasonable suspicion); *Spillner*, 173 P.3d at 506 ("The danger of 'the unbridled discretion of law enforcement officials,' also prohibits law enforcement from basing a stop solely on an officer's knowledge of a particular citizen's criminal background . . . ." (citation omitted) (quoting Delaware v. Prouse, 440 U.S. 648, 661 (1979))).

[116] *See, e.g.*, Robinson v. State, 388 So.2d 286, 290 (Fla. Dist. Ct. App. 1980) ("We hold that an officer's knowledge of a suspect's previous arrest, standing alone, is insufficient to give rise to a reasonable suspicion that a crime may have been or is being committed in order to justify a lawful investigatory stop.").

[117] *Spillner*, 173 P.3d at 506 (quoting United States v. Sandoval, 29 F.3d 537, 542-43 (10th Cir. 1994)).

there are real concerns associated with this increased access to personal information. First, the personal information can be overbroad. In *Ornelas*, for example, the officer's suspicion that a particular car was connected with a known drug dealer did not alone create a reasonable suspicion that the individual possessed drugs at the time.[118] The actions of the suspect, including staying overnight at a motel, also did not necessarily suggest drug distribution. To justify the stop, the police officer considered additional data points in his calculus, but the data themselves did not meaningfully relate to the likelihood of criminal activity at the particular time. Second, the personal information can be wrong. In *White*, for example, the tipster was wrong about some facts, including the type of narcotics police would recover from the suspect.[119] Third, the personal information can be unreliable. After *Hensley*, police-generated watchlists can be used to justify stops of individuals. As these lists are shared nationally, there is no guarantee of accuracy or any mechanism to correct mistakes.[120] Potential clerical errors, errors in judgment, and a lack of judicial oversight all create red flags for this broadening of factors included in the reasonable suspicion analysis.[121]

These concerns animate the discussion about how new data sources will affect the reasonable suspicion doctrine. Information may shape reasonable suspicion, but, as developed in the next Part, the available data can also overwhelm officers and interfere with the determination of who should be stopped for suspected criminal activity.

II. THE RISE OF BIG DATA POLICING

Big data—both as a catchphrase and a reality—is transforming the world.[122] This Part outlines the growth of big data and its potential impact

[118] *See* Ornelas v. United States, 517 U.S. 690, 692-93, 700 (1996) (remanding the case for de novo review of the district court's determination that the officer had reasonable suspicion and probable cause).

[119] Alabama v. White, 496 U.S. 325, 327 (1990).

[120] *See infra* Part IV; *see also, e.g.*, Mike McIntire, *Ensnared by Error on Growing U.S. Watch List, With No Way Out*, N.Y. TIMES, Apr. 7, 2010, at A1 (describing the difficulty of correcting mistakes on the no fly list and identifying the correct targets to include); Ellen Nakashima, *Terrorism Watch List Is Faulted for Errors*, WASH. POST, Sept. 7, 2007, at A12 (explaining problems with the accuracy of government watchlists).

[121] *See infra* Part IV; *see also* Joshua D. Wright, *The Constitutional Failure of Gang Databases*, 2 STAN. J. C.R. & C.L. 115, 118 (2005) ("[E]vidence suggests that those operating the databases are not capable of ensuring that non-gang members do not find themselves documented and trapped in the database system.").

[122] *See* Lohr, *Amid the Flood*, *supra* note 7, at BU3 ("Big Data is a shorthand label that typically means applying the tools of artificial intelligence, like machine learning, to vast new troves of data beyond that captured in standard databases."); Steve Lohr, *Sizing Up Big Data, Broadening Beyond*

on law enforcement practices. The big data revolution is just beginning, but it has already begun influencing how police identify and investigate criminal activity.

Big data will affect police officers on the streets in two primary ways. First, in conjunction with facial recognition or other biometric identification technologies,[123] unknown suspects can be known—not simply identified by name, but revealed through a web of facts involving criminal records, personal history, and past location data.[124] Vast troves of networked data can provide individualized and particularized facts from which to form suspicion.[125] Second, patterns emerging from the data will allow individuals to be identified predictively as suspects because their past actions generate suspicion about future criminal involvement.[126] Law enforcement already uses predictive policing software to predict areas of crime, but big data will soon predict actions, if not individuals.[127] The data will reveal predictive profiles to identify those believed to warrant further investigation by police.[128] In both cases, the growth of "big data" has the potential to change the reasonable suspicion calculus because more personal or predictive information about a suspect will make it easier for police to justify stopping a suspect.

the Internet, N.Y. TIMES, June 20, 2013, at F1 [hereinafter Lohr, *Sizing Up Big Data*] ("Big Data is the shorthand label for the phenomenon, which embraces technology, decision-making and public policy. . . . Big Data is a vague term, used loosely, if often, these days. But put simply, the catchall phrase means three things. First, it is a bundle of technologies. Second, it is a potential revolution in measurement. And third, it is a point of view, or philosophy, about how decisions will be—and perhaps should be—made in the future.").

[123] *See* I. Bennett Capers, *Crime, Surveillance, and Communities*, 40 FORDHAM URB. L.J. 959, 962 (2013) ("Chicago's Operation Virtual Shield includes at least 2,250 cameras, 250 of which have biometric technology."); Wayne A. Logan, *Policing Identity*, 92 B.U. L. REV. 1561, 1575 n.91 (2012) ("'Biometrics' refers either to biological or physiological characteristics usable for automatic recognition of individuals on the basis of such characteristics.").

[124] *See infra* subsection II.B.3.

[125] *See infra* Section II.B.

[126] *See infra* Section II.C.

[127] *See* Ferguson, *Predictive Policing*, *supra* note 29, at 266-67 (describing predictive policing strategies that identify the locations of potential crimes).

[128] *See* Richard Berk, *Balancing the Costs of Forecasting Errors in Parole Decisions*, 74 ALB. L. REV. 1071, 1074 (2010/2011) (discussing the use of historical data to identify future offenders); Nadya Labi, *Misfortune Teller*, ATLANTIC, Jan.–Feb. 2012, at 18-19 (discussing Professor Richard Berk's work to predict the recidivism risk of parolees in Pennsylvania).

A. *Big Data: An Introduction*

Big data refers to the accumulation and analysis of unusually large datasets.[129] It provides a shorthand term for data collection in a variety of industries and settings.[130] As described in the next few sections, this collection involves a network of sources, relying heavily on a host of consumer, social media, and law enforcement datasets, as well as more established surveillance and tracking technologies.[131]

In their book on big data, Viktor Mayer-Schönberger and Kenneth Cukier define big data in two ways. First, big data is "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value."[132] Second, they write that, "big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationships between citizens and governments, and more."[133] Jules Berman describes big data using "the three V's."[134] First, you must have "[v]olume—large amounts of data."[135] Second, you must have "[v]ariety—the data comes in different forms, including traditional databases, images, documents, and complex records."[136] Third, you must have "[v]elocity—the content of the data is constantly changing, through

[129] *See* JAMES MANYIKA ET AL., MCKINSEY GLOBAL INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY 1 (2011), *available at* http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation ("'Big data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data—i.e., we don't define big data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase.").

[130] *See* Lohr, *Amid the Flood*, *supra* note 7, at BU3 (describing the development and meaning of the term "Big Data"); *Data, Data Everywhere*, ECONOMIST (Feb. 25, 2010), http://www.economist.com/node/15557443/, *archived at* http://perma.cc/EJ3F-FDHF (noting the expansion of data collection in a variety of fields, including science, retail, and social media).

[131] *See* MANYIKA ET AL., *supra* note 129, at 5 (describing the general benefits of big data as improving transparency, facilitating experimentation, improving performance, segmenting audiences, automating decisionmaking, and enabling innovation); Lohr, *Sizing Up Big Data*, *supra* note 122, at F1 ("The bundle of technologies is partly all the old and new sources of data—Web pages, browsing habits, sensor signals, social media, GPS location data from smartphones, genomic information and surveillance videos.").

[132] *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 2 (2013).

[133] *Id.* at 6.

[134] *See* BERMAN, *supra* note 6, at *xv*, *xx*.

[135] *Id.*

[136] *Id.*

the absorption of complementary data collections, through the introduction of previously archived data or legacy collections, and from streamed data arriving from multiple sources."[137]

To understand the scope of the growth of big data, the next few sections will outline the sources, volume, and promise of big data technologies with a focus on those areas most useful for law enforcement.

B. *The Growth of Data Collection*

In many ways, the building blocks of big data are not new at all;[138] data collection has been increasing for the last few decades.[139] The growth in the volume of data collected, the ability to connect previously discrete data networks, and the analytical capabilities made possible by faster computer processors and more data storage capacity, however, are new developments.[140] These issues will be addressed in turn.

[137] *Id.*

[138] Similarly, concerns about growing data collection techniques are not new either. *See* Raymond Shih Ray Ku, *The Founders' Privacy: The Fourth Amendment and the Power of Technological Surveillance*, 86 MINN. L. REV. 1325, 1336 (2002) (discussing case law on the scope of constitutional privacy protections); Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317, 322-23 (2008) [hereinafter Slobogin, *Government Data Mining*] (discussing types of government data mining); Christopher Slobogin, *Transaction Surveillance by the Government*, 75 MISS. L.J. 139, 167-82 (2005) [hereinafter Slobogin, *Transaction Surveillance*] (outlining a potential regulatory structure for government surveillance of transactions); Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1139-40 (2002) (describing the history of data collection by all levels of government and the accompanying privacy concerns); Matthew Tokson, *Automation and the Fourth Amendment*, 96 IOWA L. REV. 581, 638-39 (2011) (worrying that Fourth Amendment law addressing letters and telephone calls is not well-suited to protect Internet communications).

[139] *See* Daniel J. Steinbock, *Designating the Dangerous: From Blacklists to Watch Lists*, 30 SEATTLE U. L. REV. 65, 69-77 (2006) (explaining the growth of government data collection to attempt to identify communists in the 1950s).

[140] *See* Joshua Gruenspecht, *"Reasonable" Grand Jury Subpoenas: Asking for Information in the Age of Big Data*, 24 HARV. J.L. & TECH. 543, 548-49 (2011) ("The increasing speed of network connections has also made possible the consolidation of computing resources and associated digital stores—a transition popularly known as the move to 'cloud computing.' When combined with the rapidly declining price of storage and the economies of scale gained from consolidating both storage and processing, the increase in network speed made it economically advantageous to store information in remote, massive data centers." (footnote omitted)); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 240 (2013) ("Big data is upon us. Over the past few years, the volume of data collected and stored by business and government organizations has exploded. The trend is driven by reduced costs of storing information and moving it around in conjunction with increased capacity to instantly analyze heaps of unstructured data using modern experimental methods, observational and longitudinal studies, and large scale simulations." (footnotes omitted)).

1. Volume of Data

The volume of collected data is growing exponentially. "The data surge just keeps rising, doubling in volume every two years. Just two days of the current global data production, from all sources—five quintillion bytes (a letter of text equals one byte)—is about equal to the amount of information created by all the world's conversations, ever . . . ."[141] This amount of information is hard to comprehend:

> [I]n 2013 the amount of stored information in the world [was] estimated to be around 1,200 exabytes, of which less than 2 percent is non-digital.
>
> There is no good way to think about what this size of data means. If it were all printed in books, they would cover the entire surface of the United States some 52 layers thick. If it were placed on CD-ROMs and stacked up, they would stretch to the moon in five separate piles. . . .
>
> Things really are speeding up. The amount of stored information grows four times faster than the world economy, while the processing power of computers grows nine times faster.[142]

Although law enforcement may not use all of this electronic data, much of it nonetheless reveals information about individuals that simply was not knowable in previous generations. As Daniel Solove has observed, "We are becoming a society of records, and these records are not held by us, but by third parties."[143]

Digital records reveal who we talk to, where we go, and what we purchase. They give insight into our hobbies, our financial status, our employment, and our criminal histories.[144] When linked together, these disparate data

[141] Lohr, *Sizing Up Big Data*, *supra* note 122, at F1.

[142] MAYER-SCHÖNBERGER & CUKIER, *supra* note 132, at 9.

[143] Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1089 (2002); *see also* James Jacobs & Tamara Crepet, *The Expanding Scope, Use, and Availability of Criminal Records*, 11 N.Y.U. J. LEGIS. & PUB. POL'Y 177, 181-82 (2008) (describing an FBI database upgrade that will give law enforcement access to "offenders' identities (name, photo, fingerprint) and the states where their rap sheets can be obtained" (footnote omitted)).

[144] *See* Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 389 (2012) ("Big data is closely linked both literally and by its scale to the massive datasets compiled by well know [sic] data aggregators such as ChoicePoint or Acxiom. Those datasets often start by aggregating large (but not "big") structured sets created by state, federal, and local governments, law enforcement, and financial institutions amongst others. Acxiom is reported to hold data on five-hundred million consumers with an average of 1500 data points per data subject." (footnotes omitted)); Janet Dean Gertz, Comment, *The Purloined Personality: Consumer Profiling in Financial Services*, 39 SAN DIEGO L. REV. 943, 944-945 (2002) (highlighting the amount of information that financial transaction data exposes, including where a person lives and works).

points can create a revealing composite of our identity, and when accessible by the government, they can serve as a valuable source of investigatory power.[145]

As the 2013 National Security Agency scandal reveals,[146] phone companies, Internet companies, and law enforcement all have the capability to store, access, analyze, and share the metadata of phone calls.[147] Metadata reveals the phone numbers contacted from a targeted phone.[148] Metadata

[145] Lior Jacob Strahilevitz, *Reputation Nation: Law in an Era of Ubiquitous Personal Information*, 102 NW. U. L. REV. 1667, 1720 (2008) ("In many instances, the government has the best access to information that decisionmakers will want to use. Criminal records, bankruptcy records, military service records, immigration and naturalization records, academic records from public schools or state-run universities, or records regarding membership in licensed professions are obvious examples."); Terry, *supra* note 144, at 389-90 ("Increasingly and of considerable importance going forward, big data comes from less structured sources including '[w]eb-browsing data trails, social network communications, sensor data and surveillance data.' Much of it is 'exhaust data,' or data created unintentionally as a byproduct of social networks, web searches, smartphones, and other online behaviors." (alteration in original) (footnotes omitted) (quoting Lohr, *Amid the Flood*, *supra* note 7)).

[146] *See* Barton Gellman & Laura Poitras, *U.S. Mines Internet Firms' Data, Documents Show*, WASH. POST, June 7, 2013, at A1; *see also* Barton Gellman & Ashkan Soltani, *NSA Taps Yahoo, Google Links*, WASH. POST, Oct. 31, 2013, at A1; Siobhan Gorman & Jennifer Valentino-DeVries, *NSA Reaches Deep into U.S. to Spy on Net*, WALL ST. J., Aug. 21, 2013, at A1; Carol D. Leonnig, Ellen Nakashima & Barton Gellman, *Judge Defends Role in Spying*, WASH. POST, June 30, 2013, at A1; Ellen Nakashima & Joby Warrick, *NSA Chief's Methods Fuel Debate on Privacy,* WASH. POST, July 15, 2013, at A1; James Risen & Laura Poitras, *N.S.A. Examines Social Networks of U.S. Citizens*, N.Y. TIMES, Sept. 29, 2013, at A1; James Risen, *Report Indicates More Extensive Cooperation by Microsoft on Surveillance*, N.Y. TIMES, July 11, 2013, at A14; Shira Ovide, *U.S. Official Releases Details of Prism Program*, WALL ST. J. (June 8, 2013, 6:28 PM), http://online.wsj.com/news/articles/SB10001424127887324299104578533802289432458, *archived at* http://perma.cc/98BU-Q9NU.

[147] *See* Ovide, *supra* note 146; *see also* Leslie Cauley, *NSA Has Massive Database of Americans' Phone Calls*, USA TODAY, May 11, 2006, at 1A; Dionne Searcey & Anne Marie Squeo, *More Phone Firms Fight Claims They Supplied Call Data to NSA*, WALL ST. J., May 17, 2006, at A3.

[148] *See* Dahlia Lithwick & Steve Vladeck, *Taking the "Meh" out of Metadata*, SLATE (Nov. 22, 2013), http://www.slate.com/articles/news_and_politics/jurisprudence/2013/11/nsa_and_metadata_how_the_government_can_spy_on_your_health_political_beliefs.html, *archived at* http://perma.cc/6PSC-5FF6 (noting also that "by analyzing the metadata of every American across a span of years, the NSA could learn almost as much about our health, our habits, our politics, and our relationships as it could by eavesdropping on our calls"); *cf.* Brian X. Chen, *Using E-Mail Data to Connect the Dots of Your Life*, N.Y. TIMES (July 5, 2013), http://bits.blogs.nytimes.com/2013/07/05/using-e-mail-data-to-connect-the-dots-of-your-life, *archived at* http://perma.cc/JHC4-Q5AN (describing how metadata programs can identify a network of contacts through linking past email contacts). Ordinary mail is also tracked in a similar way. *See* Ron Nixon, *Postal Service Is Watching Too: Outside of All Mail Is Recorded*, N.Y. TIMES, July 4, 2013, at A1 (detailing the "Mail Isolation Control and Tracking program, in which Postal Service computers photograph the exterior of every piece of paper mail that is processed in the United States—about 160 billion pieces last year").

from cell phones can reveal the location and time of a call, text, or email.[149] As one commentator explained:

> Information about where your phone has been might seem innocuous, but it can be surprisingly revealing. Location data can identify where someone sleeps, where they work, who they get a beer with, what medical professionals they visit and what political or religious gatherings they attend. And it's almost impossible to anonymize this data because . . . people are "living in habitrails," following a standardized schedule in which work and home markers are easy to discern.[150]

Location tracking through smartphone technology has become a normal part of police investigation.[151] This information is not limited to the national security context, as local law enforcement regularly requests access to phone records for ordinary criminal cases.[152] Finally, of course, the public willingly

[149] *See* Matt Richtel, *Live Tracking of Mobile Phones Prompts Court Fights on Privacy*, N.Y. TIMES, Dec. 10, 2005, at A1.

[150] Andrea Peterson, *Your Location History Is Like a Fingerprint. And Cops Can Get it Without a Warrant*, WASH. POST (July 31, 2013), http://www.washingtonpost.com/blogs/the-switch/wp/2013/07/31/your-location-history-is-like-a-fingerprint-and-cops-can-get-it-without-a-warrant, *archived at* http://perma.cc/33GH-NMFQ (quoting Jeff Jonas, IBM Fellow and Chief Scientist, IBM Entity Analytics Grp.).

[151] *Cf.* MANYIKA ET AL., *supra* note 129, at 85 ("As the number of people using mobile phones has increased, the use of cell-tower signals to triangulate the location of such devices has become increasingly common. This technology has the potential to identify the location of the owners of almost 5 billion globally."); Hayley Tsukayama, *Alarm on Hill over iPhone Location Tracking*, WASH. POST, Apr. 22, 2011, at A13; Troy Wolverton, *iSpy: Apple's iPhones Can Track Users' Movements*, SAN JOSE MERCURY NEWS (Apr. 21, 2011, 11:22 AM), http://www.mercurynews.com/ci_17893676, *archived at* http://perma.cc/9X26-TKRP.

[152] *See* Robert Block, *Requests for Corporate Data Multiply: Businesses Juggle Law-Enforcement Demands for Information About Customers, Suppliers*, WALL ST. J., (May 20, 2006, 11:59 PM), http://online.wsj.com/articles/SB114808152438358490, *archived at* http://perma.cc/KPH-3FKR; John Kelly, *Cellphone Data Spying: It's Not Just the NSA*, USA TODAY (June 13, 2014, 2:40 PM), http://www.usatoday.com/story/news/nation/2013/12/08/cellphone-data-spying-nsa-police/3902809/, *archived at* http://perma.cc/76RX-FYE3; Matt Sledge, *Cops Asked for Cell Phone Data More than 1 Million Times Last Year*, HUFFINGTON POST (Dec. 9, 2013, 3:55 PM), http://www.huffingtonpost.com/2013/12/09/cell-phone-data-requests_n_4414059.html, *archived at* http://perma.cc/48V6-C63D ("The data requests, which can be made by everyone from local cops to the FBI, are facilitated by a 1986 law called the Electronic Communications Privacy Act, which in many cases allows data content to be accessed on the simple say-so of law enforcement, without a warrant."); Bob Sullivan, *Who's Buying Cell Phone Records Online? Cops*, NBC NEWS (June 20, 2006, 11:59 AM), http://www.msnbc.msn.com/id/12534959/, *archived at* http://perma.cc/K7T4-RWA4. As will be discussed later, this information is available to police if necessary for an investigation. *See* 18 U.S.C. § 2703(f) (2012).

gives up this locational data to private companies interested in our habits and patterns.[153]

Just as people are tracked by where they go and with whom they speak, our cars and public transportation services are also tracked. Automatic license plate readers record the location of tens of thousands of cars in a growing number of cities.[154] Electronic toll collection systems record travel patterns on highways.[155] Speed cameras record travel on local roads.[156] Data recorders in our cars collect information about our driving habits, including the speed at which we drive.[157] GPS devices—using the same technology that powers our navigation systems—can track our cars.[158] Surveillance devices are being installed on public buses and subways.[159]

We reveal information about ourselves not only in the physical world, but also when we go online or use mobile applications. Some Internet and social media sites track every single click of the mouse, revealing everything

[153] *See* MANYIKA ET AL., *supra* note 129, at 86 ("A combination of navigation devices, cell-tower tracking, and smartphones accounts for the majority of personal location data today. . . . [S]martphones are a huge and fast-growing source of these data because the majority of users use applications that require their locations to be tracked."); *see also id.* at 90-91 (explaining geo-targeted advertising); *cf.* Brian X. Chen, *iPhone Tracks Your Every Move, and There's a Map for That*, WIRED (Apr. 20, 2011, 1:30 PM), http://www.wired.com/2011/04/iphone-tracks/, *archived at* http://perma.cc/Z93G-57GZ.

[154] Stephen Rushin, *The Judicial Response to Mass Police Surveillance*, 2011 U. ILL. J.L. TECH. & POL'Y 281, 286 ("ALPR [automatic license plate recognition] systems not only flag passing cars that match a criminal database, but they also record the exact time and location of *all passing cars* into a searchable database, whether or not there is any evidence of wrongdoing. This data can be kept on file indefinitely. In communities with extensive, integrated networks of ALPR cameras, this could potentially amount to mass surveillance of an entire community." (footnotes omitted)); Martin Kaste, *Police May Know Exactly Where You Were Last Tuesday*, NPR (July 17, 2013, 10:00 AM), http://www.npr.org/blogs/alltechconsidered/2013/07/16/202801282/police-may-know-exactly-where-you-were-last-tuesday, *archived at* http://perma.cc/T4NX-M7NC.

[155] *See* United States v. Jones, 132 S. Ct. 945, 963 (2012) (Alito, J., concurring in the judgment) ("[A]utomatic toll collection systems create a precise record of the movements of motorists . . . .").

[156] *See, e.g.*, Christopher Slobogin, *Public Privacy: Camera Surveillance of Public Places and the Right to Anonymity*, 72 MISS. L.J. 213, 221 (2002) (highlighting the increase in government surveillance of public places).

[157] *See, e.g.*, Associated Press, *Evidence From Black Boxes in Cars Turns Up in Courts*, FOX NEWS (June 28, 2003), http://www.foxnews.com/story/2003/06/28/evidence-from-black-boxes-in-cars-turns-up-in-courts/, *archived at* http://perma.cc/52K7-BPZP (explaining that many cars have "black boxes" that record driving behavior); Bob Gritzinger, *Under the Hood, with Big Brother*, AUTOWEEK (Nov. 7, 2004), http://autoweek.com/article/car-news/under-hood-big-brother-forget-orwells-198420-years-later-its-our-cars-are-giving-us, *archived at* http://perma.cc/9M7M-TU3W (same).

[158] *See Jones*, 132 S. Ct. at 948 (describing use of a GPS-tracking device by police).

[159] Kim Zetter, *Public Buses Across Country Quietly Adding Microphones to Record Passenger Conversations*, WIRED (Dec. 10, 2012, 4:46 PM), http://www.wired.com/2012/12/public-bus-audio-surveillance/, *archived at* http://perma.cc/8YM2-246Z.

a user does online.[160] When combined with the use of mobile communications and mobile technology, this data provides investigatory clues as to what a user does in the real world.[161]

Corporations regularly mine this online data for commercial advertising purposes.[162] In fact, both online and offline, companies create targeted consumer profiles that understand what we buy, what we do not buy,[163] and even how long we spend in particular areas of stores.[164] This information is not used just to sell things. As *Slate* reported, "for a brief period of time in

[160] In addition, the police have created software to spy on particular individuals' Internet histories. *See* Craig Timberg & Ellen Nakashima, *FBI Uses Malware to Gather Data on Suspects*, WASH. POST, Dec. 7, 2012, at A1 ("[H]igh-tech search tools, which the FBI calls 'network investigative techniques,' have been used when authorities struggle to track suspects who are adept at covering their tracks online. The most powerful FBI surveillance software can covertly download files, photographs and stored e-mails, or even gather real-time images by activating cameras connected to computers, say court documents and people familiar with this technology.").

[161] As one commentator put it:

> [T]he accumulation of a citizen's email, documents, voicemails, phone logs, records, photos, and even location by Google rivals and perhaps exceeds the data gathering capabilities of traditional law enforcement methods. . . .
>
> The synthesis of data from a user's web search history coupled with email, photos, documents, voicemails, phone logs, and location, creates a profile of an individual that serves as behavior modeling for advertisers. This same data could just as easily be disclosed to law enforcement officials for criminal profiling.

Andrew William Bagley, *Don't Be Evil: The Fourth Amendment in the Age of Google, National Security, and Digital Papers and Effects*, 21 ALB. L.J. SCI. & TECH. 153, 163-64 (2011).

[162] 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 (2008) ("A byproduct of enhanced technological capabilities is the ease with which data can be populated, aggregated, and exchanged across an increasingly diverse set of corporate interests. These corporate interests span the economy and include retailers (Sears, Hallmark), pharmaceutical companies (Pfizer), technology firms (Microsoft, IBM), banks and financial services firms (Bank One, Bank of America), and automakers (GM, Toyota). Data brokerage companies, such as Acxiom and LexisNexis repackage, augment, and sell personal data on individuals to corporate and public sector clients." (footnotes omitted)); Natasha Singer, *You for Sale: A Data Giant is Mapping, and Sharing, the Consumer Genome*, N.Y. TIMES, June 17, 2012, at BU1 (describing private-sector data mining).

[163] *See* FTC, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (2014), *available at* http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf ("[D]ata brokers hold a vast array of information on individual consumers. For example, one of the nine data brokers has 3000 data segments for nearly every U.S. consumer."). *See generally* S. COMM. ON COMMERCE, SCI., & TRANSP., MAJORITY STAFF, A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE, AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES 13-14 (2013) (describing the types of data collected by data brokers).

[164] *See* Laura Hildner, Note, *Defusing the Threat of RFID: Protecting Consumer Privacy Through Technology-Specific Legislation at the State Level*, 41 HARV. C.R.-C.L. L. REV. 133, 142 (2006) (describing radio frequency identification technology used to track customers' movements in retail stores).

2005 and 2006, the FBI, hoping to find some underground Iranian terrorist cells, . . . went through customer data collected by grocery stores in the San Francisco area searching for sales records of Middle Eastern food."[165] Because purchases can be traced to a particular point of sale, law enforcement can identify a person's location at any given point in time.[166]

Our purchases also reveal our financial resources, information that is also stored directly in numerous digital databases.[167] As anyone with credit knows, credit reports include a life's worth of financial data and life experiences.[168] The new Consumer Financial Protection Bureau has extensive data collection power over financial accounts, including personal information,[169] but this pales in comparison to the data held by private information aggregators who have developed lucrative business models around the collection and aggregation of personal information.[170]

> There are information aggregation businesses in the private sector that already combine personal data from thousands of private-sector sources and public records. ChoicePoint, Acxiom, LexisNexis, the three national credit bureaus, and dozens of other companies maintain rich repositories of information about virtually every adult in the country. These records are updated daily by a steady stream of incoming data. They provide a one-stop-shop for the government when it wants access to personal data, and most of the government's data mining initiatives depend on access to those data.[171]

[165] Evgeny Morozov, *Connecting the Dots, Missing the Story*, SLATE (June 24, 2013, 7:45 AM), http://www.slate.com/articles/technology/future_tense/2013/06/with_big_data_surveillance_the_government_doesn_t_need_to_know_why_anymore.html, *archived at* http://perma.cc/FuF4-DDRW.

[166] MANYIKA ET AL., *supra* note 129, at 85 ("Globally in 2008, there were 90 billion to 100 billion such transactions off line linkable to [point of sale] devices. Law enforcement investigations regularly use such data to establish physical location.").

[167] *Cf.* Sam Kamin, *The Private Is Public: The Relevance of Private Actors in Defining the Fourth Amendment*, 46 B.C. L. REV. 83, 125-27 (2004) (discussing databases of information on consumers that retailers compile).

[168] *See* Chris Jay Hoofnagle, *Big Brother's Little 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, 600-07 (2004) (describing the types of data compiled by different information firms, including the credit bureau Experian).

[169] *See* Carter Dougherty, *Consumer Bureau Chief Defends Big-Data Program*, BOS. GLOBE, Apr. 24, 2013, at B10.

[170] Jon D. Michaels, *All the President's Spies: Private-Public Intelligence Partnerships in the War on Terror*, 96 CALIF. L. REV. 901, 902 (2008) ("[P]rivate organizations can at times obtain and share information more easily and under fewer legal restrictions than the government can when it collects similar information on its own.").

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

While federal, state, and local laws limit direct government access to financial records without some legal process,[172] the government may indirectly access this same information through data aggregating services almost without restriction.[173]

Finally, police, of course, access law enforcement records of past convictions, arrests, and information related to those contacts. Most officers have access to the National Crime Information Center (NCIC), a computerized database of criminal justice information.[174] According to internal FBI reports, users searched the NCIC database 2.7 billion times in 2011 and the database had 11.7 million active records.[175] Once police have accessed the NCIC system, they can pull up physical characteristics or addresses and query the database to determine whether observed suspects live in an area or whether they match a description of a wanted suspect.[176]

2. Networked Data

The investigatory utility of standalone databases improves when law enforcement agencies and private companies connect those databases and aggregate their data. Indeed, linking traditional criminal justice data with private data provides a wealth of insights about a person.[177] In recent years,

[172] *See, e.g.,* Eric Lichtblau & James Risen, *Bank Data Sifted in Secret by U.S. to Block Terror*, N.Y. TIMES, June 23, 2006, at A1; Eric Lichtblau, *F.B.I.'s Reach Into Records Is Set To Grow*, N.Y. TIMES, Nov. 12, 2003, at A12 (characterizing the government's ability to gather financial data on individuals as at least moderately constrained); Josh Meyer & Greg Miller, *U.S. Secretly Tracks Global Bank Data*, L.A. TIMES, June 23, 2006, at A1.

[173] *See* Joshua L. Simmons, Note, *Buying You: The Government's Use of Fourth-Parties to Launder Data About 'The People,'* 2009 COLUM. BUS. L. REV. 950, 951-52, 990-99 (reporting that the government turns to private companies to provide information that it would be restricted from collecting on its own); Pratap Chatterjee, *The Data Hackers*, NATION (Oct. 8, 2013), http://www.thenation.com/article/176542/data-hackers, *archived at* http://perma.cc/DRG8-62XQ (reporting that private companies sell data to the government for law enforcement purposes).

[174] *Cf. National Crime Information Center*, FBI, *supra* note 11.

[175] *The CJIS Division Turns 20*, CJIS LINK (Fed. Bureau of Investigation/Criminal Justice Info. Servs. Div.), Mar. 2012, at 3.

[176] *Cf.* Kline, *supra* note 162, at 451 ("An example of a state-level database initiative is the Multi-State Anti-Terrorism Information Exchange ('MATRIX'), a law enforcement database that combines data from private and public sources to create a searchable database to assist in police investigations.").

[177] Slobogin, *Transaction Surveillance*, *supra* note 138, at 145 ("[A]dvances in data warehousing and data exchange technology in the financial sector allow very easy access to a virtual cornucopia of transaction-related information that can reveal, among other things, 'what products or services you buy; what charities, political causes, or religious organizations you contribute to; . . . where, with whom, and when you travel; how you spend your leisure time; . . . whether you have unusual or dangerous hobbies; and even whether you participate in certain felonious activities.'" (quoting Gertz, *supra* note 144, at 944-45)); Solove, *supra* note 138; Daniel J. Solove, *Data Mining and the Security–Liberty Debate*, 75 U. CHI. L. REV. 343, 344 (2008) [hereinafter Solove, *Security–*

the federal government created two such networked database programs, but eventually discontinued their use due to public concerns about privacy.

The first database, the Multi-State Anti-Terrorism Information Exchange Program (MATRIX), included a networked database that allowed police officers to check a broad range of information with one search, including criminal history, credit information, driver's license information, vehicle registration, arrests, utility connections, UCC filings, concealed weapons permits, FAA aircraft and pilots licenses, hunting and fishing licenses, professional licenses, and voter registration records.[178] As the Electronic Privacy Information Center argued in its amicus brief in *Hiibel v. Sixth Judicial District Court of Nevada*, a police officer using the MATRIX system could develop an entire profile of a suspect simply by running a name in the database during a routine encounter.[179] According to the Department of Homeland Security's Privacy Office in its review of the program,

> only 2.6% of the cases investigated over the course of the MATRIX pilot project were related to terrorism. In fact, the MATRIX project was predominantly used to investigate fraud, robbery, and other crimes, including assault, homicide and narcotics cases, underscoring the value of the program as a tool for traditional law enforcement.[180]

The second database, the even more Orwellian-sounding Total Information Awareness System, was designed by the Department of Defense to fight terrorism by linking data sources into one searchable national information collection center.[181] Renamed the Terrorism Information Awareness

Liberty Debate] (introducing government data mining programs); Omer Tene, *What Google Knows: Privacy and Internet Search Engines*, 2008 UTAH L. REV. 1433, 1458 (explaining that search query logs aggregate vast amounts of personal information).

[178] Brief of Amici Curiae Electronic Privacy Information Center (EPIC) et al. at 12-13, Hiibel v. Sixth Judicial Dist. Court of Nev., 542 U.S. 177 (2004) (No. 03-5554); *see also* U.S. DEP'T OF HOMELAND SEC., REPORT TO THE PUBLIC CONCERNING THE MULTISTATE ANTI-TERRORISM INFORMATION EXCHANGE (MATRIX) PILOT PROJECT 2-4 (2006), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy-matrix-122006.pdf (describing privacy concerns that doomed the project); Katie Stenman, Comment, *State Government Information Collection: The Shutdown of the MATRIX Program, REAL ID, and DNA Collection*, 2 INFO. SOC'Y J.L. & POL'Y 547, 549-50 (2006) (describing the MATRIX database).

[179] Brief of Amici Curiae, *supra* note 178, at 30.

[180] U.S. DEP'T OF HOMELAND SEC., *supra* note 178, at 2.

[181] *See* Douglas A. Fretty, Comment, *Face-Recognition Surveillance: A Moment of Truth for Fourth Amendment Rights in Public Places*, 16 VA. J.L. & TECH. 430, 435-36 (2011) ("Called TIA (originally 'Total Information Awareness' but redubbed 'Terrorism Information Awareness' to avoid an overtly Orwellian moniker), the program included a HumanID component, intended to 'identify humans using a combination of biometric modes at distances up to 500 feet.'" (footnotes omitted)).

system,[182] this program had several components all designed to aggregate available information for predictive surveillance purposes.[183] The Department of Defense reportedly designed the program to "connect the dots" in an attempt to avoid repeating the missed opportunities to intervene before the terrorist attacks of September 11, 2001.[184] Among other things, the program sought to create predictive "risk profiles" for particular citizens based on the available data.[185]

While these two programs (and others) were canceled over privacy concerns,[186] law enforcement and private companies have embraced the idea of networking and sharing personal information. First, the Federal Bureau of Investigation's Criminal Justice Information Services Division coordinates access to databases that include public and private sources.[187] FBI agents

[182] *Id.*

[183] *See* Slobogin, *Government Data Mining*, *supra* note 138, at 318 ("Beginning soon after the passing of TIA, it spent at least $40 million developing a program called ADVISE (for Analysis, Dissemination, Visualization, Insight, and Semantic Enhancement), which was designed 'to troll a vast sea of information, including audio and visual, and extract suspicious people, places and other elements based on their links and behavioral patterns.'" (quoting Ellen Nakashima & Alec Klein, *New Profiling Program Raises Privacy Concerns*, WASH. POST, Feb. 28, 2007, at B1)).

[184] *See* K.A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 1, 3 n.3 (2003) (remarking that the Joint Inquiry Into the Intelligence Community Activities Before and After the Terrorist Attacks of September 11, 2001, H.R. REP. NO. 107-792, S. REP. NO. 107-351 (2002), "refers at least ten times to the intelligence communit[y's] failure to 'connect the dots'"); *see also* Fretty, *supra* note 181.

[185] *See* Solove, *Security–Liberty Debate*, *supra* note 177, at 343 ("Under the TIA program, the government would assemble a massive database consisting of financial, educational, health, and other information on US citizens, which would later be analyzed to single out people matching a terrorist profile. According to [Admiral John] Poindexter, '[t]he only way to detect . . . terrorists is to look for patterns of activity that are based on observations from past terrorist attacks as well as estimates about how terrorists will adapt to our measures to avoid detection.'" (quoting John M. Poindexter, Op-Ed, *Finding the Face of Terror in Data*, N.Y. TIMES, Sept. 10, 2003, at A25)); Jeffrey Rosen, *Total Information Awareness*, N.Y. TIMES, Dec. 12, 2002 (Magazine), at 128, 129 ("In addition to analyzing financial, educational, travel and medical records, as well as criminal and other governmental records, the T.I.A. program could include the development of technologies to create risk profiles for millions of visitors and American citizens in its quest for suspicious patterns of behavior.").

[186] *See* U.S. DEP'T OF HOMELAND SEC., *supra* note 178, at 4 (analyzing the cancellation of MATRIX); Solove, *Security–Liberty Debate*, *supra* note 177, at 1 (reporting the cancellation of TIA, but suggesting that it was merely replaced with similar programs).

[187] *See* Cate, *supra* note 171, at 442-43 ("The Federal Bureau of Investigation ('FBI') maintains extensive databases in its Criminal Justice Information Services Division ('CJISD') that collect data from, and supply data to, a wide array of public- and private-sector entities."); Hoofnagle, *supra* note 168, at 599-600 (describing the murky but robust relationship between government agencies and private data brokers); Rushin, *supra* note 154, at 292 ("In 2003, the Department of Justice began the National Criminal Intelligence Sharing Plan (NCISP), which is designed at improving the sharing of criminal intelligence data.").

and analysts regularly access these databases which contain hundreds of millions of records.[188]

Law enforcement has benefited and continues to benefit from the growth of private surveillance collection services.[189] For example, the technology that ran the Total Information Awareness Program is now owned by LexisNexis, a private company.[190] As one expert has written, "[T]he private sector is developing domain specific technologies (that is, applications developed specifically for law enforcement purposes) to aggregate and mine data using both link analysis and pattern-matching in criminal investigations and these technologies are already being adopted and employed in a variety of law enforcement environments."[191] Police, thus, can and do request information from third-party data sources, including commercial data aggregators, Google, phone companies, and social and financial networks.[192] This creates the potential to replicate in the private

[188] *See* Cate, *supra* note 171, at 444 ("The FBI aggregates data from multiple databases into its Investigative Data Warehouse ('IDW'). According to press briefings given by the FBI in 2006, the IDW contains more than 659 million records, which come from 50 FBI and outside government agency sources. The system's data mining tools are so sophisticated that they can handle many variations in names and other data, including up to twenty-nine variants of birth dates. The 13,000 agents and analysts who use the system average one million queries a month." (footnotes omitted)); Slobogin, *Government Data Mining*, *supra* note 138, at 319-20 ("The DOJ, through the FBI, has been collecting telephone logs, banking records, and other personal information regarding thousands of Americans not only in connection with counterterrorism efforts, but also in furtherance of ordinary law enforcement." (footnote omitted)).

[189] *See* Simmons, *supra* note 173, at 951-52 ("Your information is for sale, and the government is buying it at alarming rates. The CIA, FBI, Justice Department, Defense Department, and other government agencies are, at this very moment, turning to a group of companies to provide them with information that these companies can gather without the restrictions that bind government intelligence agencies.").

[190] Joseph T. Thai, *Is Data Mining Ever a Search Under Justice Stevens's Fourth Amendment?*, 74 FORDHAM L. REV. 1731, 1739-40 (2006) ("MATRIX allowed law enforcement to search through 'billions of records from disparate datasets' from participating states as well as 'commercially available data sources.' The database itself actually was developed and maintained by Seisint, Inc., based on its Accurint service that LexisNexis later acquired." (footnote omitted)).

[191] Taipale, *supra* note 184, at 15 (footnote omitted); *see also id.* at 14-15 ("The notion that powerful analytical tools developed for commercial and scientific application will not eventually be used for terrorism prevention (or, for that matter, general law enforcement purposes) seems unrealistic, particularly since these technologies are already being used in a wide variety of law enforcement contexts." (footnote omitted)).

[192] *See, e.g.*, Editorial, *The End of Privacy*?, N.Y. TIMES, July 15, 2012, at SR10 (describing increased opportunities for surveillance in a digital era); *see also* Ira S. Rubinstein, Ronald D. Lee & Paul M. Schwartz, *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*, 75 U. CHI. L. REV. 261, 271-73 (2008) (same); Taipale, *supra* note 184, at 21 ("In the context of law enforcement, data mining is no more than the computational automation of traditional investigative skills—that is, the intelligent analysis of myriad 'clues' in order to develop a theory of the case."); Terry, *supra* note 144, at 391 ("Big data is creating a *private* surveillance model that will exceed law enforcement tracking of individuals using Internet and cell phone

sector much of what was envisioned in the original government surveillance projects that were considered threatening to Americans' privacy.[193]

Finally, because commercial entities—rather than the government—own these "fourth-party" records, they avoid many of the constitutional and statutory protections that might ensure privacy of these records.[194] "Today, data aggregators are able to cross-index various sources of information to produce incredibly extensive—and invasive—lists for practically any purpose. For example, many can 'provide lists of people who take Prozac for depression, believe in the Bible, gamble online, or buy sex toys.'"[195]

data."); Lee Tien, *Privacy, Technology and Data Mining*, 30 OHIO N.U. L. REV. 389, 390 (2004) (highlighting the vast amount of data held by private firms and suggesting that the government would like to use it for law enforcement); Glenn R. Simpson, *Big Brother-in-Law: If the FBI Hopes To Get the Goods on You, It May Ask ChoicePoint*, WALL ST. J., Apr. 13, 2001, at A1 (reporting on the ease with which the government can access information on individuals through commercial data purchases); Andy Greenberg, *U.S. Government Requests for Google Users' Private Data Jump 37% in One Year*, FORBES (June 17, 2012, at 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, *archived at* http://perma.cc/HQP2-PNPM (reporting on government requests for data held by Google); Declan McCullagh, *Feds Push for Tracking Cell Phones*, CNET (Feb. 11, 2010, 4:00 AM), http://www.cnet.com/news/feds-push-for-tracking-cell-phones, *archived at* http://perma.cc/4TXW-ZX64 (discussing government efforts to use cell phone data for law enforcement).

[193] ChoicePoint itself is said to have "14 billion records on individuals and businesses that can be used for tasks like pre-employment screening of job candidates." Kline, *supra* note 162, at 448 ("Electronically available personal data culled from public and private records forms the backbone of the multi-billion dollar database-marketing industry. Data brokers and their customers collect and trade massive amounts of digitized personal data on most Americans through database-driven information markets."); *see also* Strahilevitz, *supra* note 145, at 1670 ("One of the most significant developments in the industrialized world during the last decade has been the increased availability of information about individuals. Personal information that was once obscure can be revealed almost instantaneously via a Google search.").

[194] *See* Simmons, *supra* note 173, at 976 ("There is no provision . . . preventing [a] financial institution from disclosing . . . information to a fourth-party, who could then pass it on to the government."); *see also* Michaels, *supra* note 170, at 902.

[195] Simmons, *supra* note 173, at 990-91 (quoting Paul Magnusson, *They're Watching You*, BUS. WK., Jan. 24, 2005, at 22); *see also* JULIA ANGWIN, DRAGNET NATION: A QUEST FOR PRIVACY, SECURITY, AND FREEDOM IN A WORLD OF RELENTLESS SURVEILLANCE 3 (2014) (questioning whether companies ought to "scoop up information about people's mental health"); Elspeth A. Brotherton, Comment, *Big Brother Gets a Makeover: Behavioral Targeting and the Third-Party Doctrine*, 61 EMORY L.J. 555, 563 (2012) ("For example, 'supermassive databases'—like those made available by companies such as LexisNexis—offer *billions* of records about individuals aggregated from public and private records. Thus, a user's profile could reflect vast quantities of highly sensitive personal information, including the user's 'demographics, family information, and credit history.'" (footnote omitted)); Julia Angwin, *The Web's New Gold Mine: Your Secrets*, WALL ST. J., July 31, 2010, at W1 ("Microsoft . . . had a prediction of . . . age, ZIP code[,] . . . gender[,] . . . income, marital status, presence of children and home ownership."); John Markoff, *You're Leaving a Digital Trail. Should You Care?*, N.Y. TIMES, Nov. 30, 2008, at BU1 (highlighting the abundance of smartphone-generated data); Robert O'Harrow Jr., *In Age of Security, Firm Mines Wealth of Personal*

Similarly, the aggregators can tailor searches to identify those allegedly engaged in illicit activities, who otherwise would avoid suspicion. Such tools are increasingly useful to generate personalized, individualized information about a suspect, and remain largely unregulated.

3. Identifiable Data

To solve crimes, law enforcement must not only collect information, but also identify and link individuals to their accumulated data. In short, data must be connected with identifiable human beings.

Facial recognition software, biometric identification technologies, and mobile communication make it easier to identify unknown suspects and access data associated with these suspects.[196] Today, facial recognition software can identify a suspect by comparing the observed suspect's face to a database of stored faces.[197] As sources of photographs proliferate, and law enforcement databases link these sources together, the utility and ease of the technology will expand rapidly.[198] In more technologically advanced

Data, WASH. POST, Jan. 20, 2005, at A1 (reporting on ChoicePoint's growth in private- and public-sector clients).

[196] *See* Aliya Sternstein, *FBI Seeks Video Recognition Technology to Automatically ID Suspects*, NEXTGOV (Nov. 4, 2013), http://www.nextgov.com/emerging-tech/2013/11/fbi-seeks-video-recognition-technology-automatically-id-suspects/73168/, *archived at* http://perma.cc/X5H2-CWR3 ("The FBI is weighing the use of video recognition technology to quickly identify suspects, even if all the camera has captured is a perpetrator's limp or fraying blue baseball cap. Think of it as automated police lineups for the YouTube generation."); *see also* Marc Jonathan Blitz, *Video Surveillance and the Constitution of Public Space: Fitting the Fourth Amendment to a World That Tracks Image and Identity*, 82 TEX. L. REV. 1349, 1354 (2004) (highlighting identification technologies); John J. Brogan, *Facing the Music: The Dubious Constitutionality of Facial Recognition Technology*, 25 HASTINGS COMM. & ENT. L.J. 65, 81 (2002) (same).

[197] *See* Rushin, *supra* note 154, at 288 ("During the Super Bowl in 2001, FaceTrac technology was used to digitally scan 128 points on the face of each fan entering Raymond James Stadium in Tampa, Florida. This information was then compared to Federal Bureau of Investigations [sic] databases. In total, the technology was able to identify nineteen suspected criminals. Similar technology has been employed in major cities across the country including Boston, Tampa, Providence, Kansas City, and Washington, D.C." (footnotes omitted)).

[198] Laura K. Donohue, *Technological Leap, Statutory Gap, and Constitutional Abyss: Remote Biometric Identification Comes of Age*, 97 MINN. L. REV. 407, 447-48 (2012) ("[T]he Interstate Photo System (IPS) . . . draws heavily on FRT and data mining technologies—and the database on which it is built is rapidly growing. As of 2009, [the database] included more than 6.75 million photos. By February 2012, this number had increased to more than 114.5 million photos." (footnotes omitted)); Margaret Hu, *Biometric ID Cybersurveillance,* 88 IND. L.J. 1475, 1521 (2013) ("With a universal biometric database and 'cardless' national ID system, such as a biometric E-Verify system, or biometric national ID card—e.g., digitalized and multimodal biometric driver's license, Social Security Card, or passport—federal, state, and local law enforcement could scan biometric data or request to see a digitalized biometric ID for a wide range of reasons, including routine traffic stops.").

jurisdictions, mobile handheld devices can match faces to a central database.[199] Soon the technology will complete searches in real time by allowing police to scan multiple faces along a street.[200] As more police use portable computers linked to these photograph databases, the ability of law enforcement to scan and analyze faces to identify suspects will become more common.[201]

The scale of available photo databases demonstrates the power of networked data. As reported by *The Washington Post*, "The FBI's own facial-recognition database has about 15 million criminal mug shots. Bureau officials are pushing to expand that by tens of millions more by encouraging states to upload their criminal justice photos into the national system."[202] Many states have complied or are complying with this request, and some have created their own systems. The Pinellas County Sheriff's Office in Florida, for example, has built one of the country's most advanced

[199] *See, e.g.*, Sabrina A. Lochner, Note, *Saving Face: Regulating Law Enforcement's Use of Mobile Facial Recognition Technology & Iris Scans*, 55 ARIZ. L. REV. 201, 206 (2013) ("[T]he American military began using a multi-modal device called Handheld Interagency Identity Detection Equipment ('HIIDE') in 2007. This allowed soldiers to take facial pictures, iris scans, and fingerprints in the field and compare the gathered information to a database; the comparison let soldiers see if the person being scanned was on a watch list and allowed the soldiers to determine the person's identity.").

[200] *See* Craig Timberg & Ellen Nakashima, *Photo-ID Databases Become Troves for Police*, WASH. POST, June 17, 2013, at A1 ("But research efforts are focused on pushing the software to the point where it can reliably produce the names of people in the time it takes them to walk by a video camera. This already works in controlled, well-lit settings when the database of potential matches is relatively small. Most experts expect those limitations to be surmounted over the next few years."); David Goldman, *Real-Time Face Recognition Comes to Your iPhone Camera*, CNN MONEY (Mar. 12, 2012, 11:13 AM), http://money.cnn.com/2012/03/12/technology/iPhone-face-recognition/, *archived at* http://perma.cc/J3WH-2Z3V; Zach Howard, *Police to Begin iPhone Iris Scans Amid Privacy Concerns*, REUTERS, July 20, 2011, *available at* http://www.reuters.com/article/2011/07/20/us-crime-identification-iris-idUSTRE76J4A120110720.

[201] *See* Lochner, *supra* note 199, at 202 ("Beginning in April 2012, more than 50 law enforcement agencies across the United States began using a mobile device to identify people through facial recognition technology ('FRT'), iris scans, and fingerprints." (footnote omitted)); *id.* at 205 ("Using FRT, police can determine someone's identity by running a photo of that person's face through a database. The computer program matches the unidentified face with a picture, name, and criminal record of someone already in the database." (footnote omitted)); *see also* Fretty, *supra* note 181, at 435 ("[C]ities are embracing FRT to monitor their citizens on a daily, more mundane basis. Many municipalities, including Los Angeles and New York City, have equipped police officers with facial scanners that determine whether a suspect has a criminal record, while others install the technology on stationary street cameras." (footnote omitted)).

[202] Timberg & Nakashima, *supra* note 200, at A1; *see also* Ryan Gallagher, *FBI to Give Facial Recognition Software to Law-Enforcement Agencies*, SLATE (Aug. 23, 2012, 5:08 PM), http://www.slate.com/blogs/future_tense/2012/08/23/universal_face_workstation_fbi_to_give_facial_recognition_software_to_law_enforcement_.html, *archived at* http://perma.cc/G9VR-NMBF; Sara Reardon, *FBI Launches $1 Billion Face Recognition Project*, NEWSCIENTIST (Sept. 7, 2012), http://www.newscientist.com/article/mg21528804.200-fbi-launches-1-billion-facerecognition-project.html, *archived at* http://perma.cc/S3Q4-4WUM.

facial-recognition programs.[203] "The faces of more than 120 million people are in searchable photo databases that state officials assembled to prevent driver's license fraud but that increasingly are used by police to identify suspects, accomplices and even innocent bystanders in a wide range of criminal investigations."[204] "Pennsylvania's Justice Network, which has allowed police anywhere in the state to compare a facial image with mug-shot databases, has become a key investigative tool . . . and last month it added access to 34 million driver's-license photos."[205] These examples represent only the beginning as twenty-six states now allow local law enforcement to access driver's license photographs for facial recognition purposes.[206] In addition, facial recognition programs can easily search a wealth of personal photographs uploaded online.[207]

Facial recognition is but one technology used to identify suspects on the streets. Law enforcement can also use biometric identification technologies[208] that look to irises, tattoos, scars, face-shape, and even the habitual

[203] Timberg & Nakashima, *supra* note 200, at A1.

[204] *Id.*

[205] *Id.*

[206] *Id.* ("Thirty-seven states now use facial-recognition technology in their driver's-license registries . . . . At least 26 of those allow state, local or federal law enforcement agencies to search—or request searches—of photo databases in an attempt to learn the identities of people considered relevant to investigations."); *see also id.* ("The increasingly widespread deployment of the technology in the United States has helped police find murderers, bank robbers and drug dealers, many of whom leave behind images on surveillance videos or social-media sites that can be compared against official photo databases. . . . [L]aw enforcement use of such facial searches is blurring the traditional boundaries between criminal and non-criminal databases, putting images of people never arrested in what amount to perpetual digital lineups. The most advanced systems allow police to run searches from laptop computers in their patrol cars and offer access to the FBI and other federal authorities.").

[207] *See* Emily Steel, *A Face Launches 1,000 Apps*, WALL ST. J., Aug. 5, 2011, at B5 (reviewing the proliferation of social media applications that use facial recognition); Richard Lardner, *Your New Facebook 'Friend' May Be the FBI*, NBC NEWS (Mar. 16, 2010, 10:54 AM), http://www.nbcnews.com/id/35890739/ns/technology_and_science-security/t/your-newfacebook-friend-may-be-fbi/, *archived at* http://perma.cc/KQJ2-9MT3 (reporting on use of social media by law enforcement); *see also* Dino Grandoni, *Facebook's New "DeepFace" Program Is Just As Creepy As It Sounds*, HUFFINGTON POST, http://www.huffingtonpost.com/2014/03/18/facebook-deepface-facial-recognition_n_4985925.html (last updated Mar. 25, 2014, 2:59 PM), *archived at* http://perma.cc/79TP-MEJR ("Facebook owns the world's largest photo library, and it now has the technology to match almost all the faces within it. Yes, even the ones you don't tag. Facebook announced . . . that it has developed a program called 'DeepFace,' which researchers say can determine whether two photographed faces are of the same person with 97.25 percent accuracy.").

[208] Daniel J. Steinbock, *National Identity Cards: Fourth and Fifth Amendment Issues*, 56 FLA. L. REV. 697, 704-05 (2004) ("Biometrics are identification techniques based on some unique, physiological, and difficult-to-alienate characteristic. Current forms of identification often rely on relatively primitive biometrics such as skin, hair and eye color, physical markings, gender, and facial hair. These characteristics are often portrayed in a photograph or list of physical characteristics, such as those used on a driver's license." (footnote omitted)).

manner in which people walk.[209] The goal: a comprehensive remote scanning technology that would allow instant identification of suspects through a quick search of a massive database.[210] One existing program, dubbed MORIS, already allows an ordinary iPhone user to scan an iris and compare it to a database of biometric identifiers.[211] With MORIS,

> [p]olice can take a picture of the subject's face from up to [five] feet away and conduct an iris scan from up to [six] inches from the person's eye. The device matches photographs against a national criminal records database that is managed by Biometric Intelligence and Identification Technologies ('BI2 Technologies'), the private company that designed MORIS.[212]

Other techniques will also be developed that may allow similar scanning through other surveillance techniques.[213]

These growing mobile technologies not only allow law enforcement to identify previously unknown suspects but also provide other networked personal information about those suspects.[214] As one state law enforcement

[209] Ellen Nakashima, *FBI Prepares Vast Database of Biometrics*, WASH. POST, Dec. 22, 2007, at A1; *Image-Based Matching Technology Offers Identification and Intelligence Prospects*, CJIS LINK (FBI/Criminal Justice Info. Servs. Div), Dec. 2012, at 4 ("In 2014, investigators will be able to query the NGI [(Next Generation Identification)] with descriptive data about tattoos to find images of potential matches of [scars, marks, and tattoos] associated with individuals' records.").

[210] Wendy Koch, *Iris Scans Let Law Enforcement Keep an Eye on Criminals*, USA TODAY, Dec. 5, 2007, at A1; Howard, *supra* note 200.

[211] Lochner, *supra* note 199, at 207 ("MORIS attaches to an iPhone and allows law enforcement officers to search facial, iris, and fingerprint databases while they are in the field.")

[212] *Id.* at 208 (footnotes omitted); *see also* Donohue, *supra* note 198, at 461-62 ("[T]he Mobile Offender Recognition and Information System, known as MORIS, incorporates FRT, iris scans, and fingerprinting. Police officers equipped with the device can take a picture of a person's face from a distance of two to five feet away, which is then analyzed according to 130 distinguishing points. This information can then be compared to existing databases." (footnotes omitted)); Emily Steel, *How a New Police Tool for Face Recognition Works*, WALL ST. J. (July 13, 2011, 7:56 AM), http://blogs.wsj.com/digits/2011/07/13/how-a-new-police-tool-for-face-recognition-works/, *archived at* http://perma.cc/TXS-8DHQ (noting the MORIS device's usefulness in identifying individuals who are not carrying forms of identification).

[213] *See, e.g.*, Emily Steel & Julia Angwin, *Device Raises Fear of Facial Profiling*, WALL ST. J., July 13, 2011, at A1; Tovia Smith, *New Police Scanner Raises "Facial Profiling" Concerns*, NPR (July 27, 2011, 9:58 PM), http://www.npr.org/2011/08/11/138769662/new-police-scanner-raises-facial-profiling-concerns, *archived at* http://perma.cc/R5HT-Y342; *see also* Noah Shachtman, *Army Tracking Plan: Drones that Never Forget a Face*, WIRED (Sept. 28, 2011, 6:30 AM), http://www.wired.com/2011/09/drones-never-forget-a-face/, *archived at* http://perma.cc/FP49-4KTZ.

[214] *See* Donohue, *supra* note 198, at 412-13 ("The Federal Bureau of Investigation (FBI), for example, is currently developing what it calls Next Generation Identification (NGI). One of its components, the Interstate Photo System, allows law enforcement to submit still images or video surveillance feeds obtained from any public or private source. The system is designed to store this data and, using FRT, to identify individuals, pairing images with biographic information. NGI

officer remarked, "I can call up everything about you, your pictures and pictures of your neighbors."[215] In addition, these technologies will allow law enforcement to identify targeted populations easily, such as gang members or suspects identified on "most wanted lists."[216] For example, police in Lincoln, Nebraska, carry a mobile application called "P3I" (Proactive Police Patrol Information) that displays the location of suspected gang members, registered sex offenders, people with outstanding warrants, parolees, and criminal incident reports.[217] The result is that big data technology can provide vastly more identifying information to help determine reasonable suspicion on the streets.

C. *Predictive Data*

Big data also promises another change in law enforcement techniques. The expansive collection of data allows for more sophisticated analysis that might reveal previously unknown patterns of criminal activity.[218]

also uses biographic information to search its Repository for Individuals of Special Concern (RISC)." (footnotes omitted)).

[215] Simmons, *supra* note 173, at 952 n.1.

[216] *See* Molly Bruder, Comment, *Say Cheese! Examining the Constitutionality of Photostops*, 57 AM. U. L. REV. 1693, 1697 (2008) ("Increasingly, police departments and law enforcement agencies are using gang databases to combat gang violence. These databases contain personal information about suspected gang members, including gang allegiance, street name, address, physical description, identifying marks, tattoos, and photographs." (footnote omitted)); Jim Adams, *Officers Share Names to Battle Gangs*, STAR TRIB. (Minneapolis), Feb. 24, 1998, at B1; Editorial, *"GangNet" Bears Watching*, DENVER POST, Sept. 28, 2002, at B23; Ryan Lizza, *The Year in Ideas: Ghetto Profiling*, N.Y. TIMES, Dec. 15, 2002 (Magazine), at 94, 94-95 (describing a profiling technique in which police target crime-plagued neighborhoods to build a database of potential suspects); *see also* Hong H. Tieu, *Picturing the Asian Gang Member Among Us*, 11 ASIAN PAC. AM. L.J. 41, 44-45 (2006) (reporting that California's "CalGang" database is the largest gang database in the nation and contains photographs of suspected gang members who have been detained—although not necessarily arrested—by local police departments); De Tran & Iris Yokoi, *O.C. Asians Say Police Photos Are Harassment: Dispute: Fountain Valley's "Mug" Shots Unfairly Stereotype Youths as Gang Members, Complainants Say*, L.A. TIMES (Nov. 15, 1992), http://articles.latimes.com/1992-11-15/news/mn-1093_1_fountain-valley-police-department, *archived at* http://perma.cc/CQK3-3NC6 (reporting that Asian youths allege that they are unfairly branded as gang members as police take their photos for the gang database).

[217] Tom Casady, *P3i Lincoln Police Department*, YOUTUBE (July 28, 2011), https://www.youtube.com/watch?v=7HpQwkAcU24&f; Zach Pluhacek, *Lincoln Cops' App Up for Download*, LINCOLN J. STAR (Aug. 24, 2011, 9:00 PM), http://journalstar.com/news/local/crime-and-courts/lincoln-cops-app-up-for-download/article_6a2ae7c2-4597-51e0-a5b3-eae4069a587a.html, *archived at* http://perma.cc/5XJ5-AJW3; *see also* Zach Pluhacek, *Finding Crooks? "There's an App for That,"* LINCOLN J. STAR (Oct. 15, 2010, 6:45 AM), http://journalstar.com/news/local/crime-and-courts/finding-crooks-there-s-an-app-for-that/article_d3f33ae6-d7ea-11df-b5d6-001cc4c03286.html, *archived at* http://perma.cc/A7R3-JCCG.

[218] *See* Steve Lohr, *The Age of Big Data*, N.Y. TIMES, Feb. 11, 2012, at SR1 ("Police departments across the country, led by New York's, use computerized mapping and analysis of variables

Predictive policing technologies are already in use in major cities such as Los Angeles and Seattle.[219] While current technologies focus primarily on expected "places" of criminal activity, they also predict patterns of criminal actions. Computer programs can analyze these patterns, and police accordingly could use these patterns to stop unknown suspects whose actions fit the predicted activity.[220] Available technologies provide different levels of sophistication, but the underlying theory that crime patterns can be identified, analyzed, and predicted is well established.[221] Traditional hot-spot mapping,[222] COMPSTAT systems,[223] and more modern technologies like predictive policing all rely on data analysis to track crime patterns.[224] The logic behind these technologies is that certain environmental vulnerabilities exist to encourage crime at a particular location.[225] Repeated observation of

like historical arrest patterns, paydays, sporting events, rainfall and holidays to try to predict likely crime 'hot spots' and deploy officers there in advance.").

[219] *See* Kahn, *supra* note 24 (describing predictive policing in Los Angeles); Rubin, *supra* note 24 (same); *see also* Martin Kaste, *Can Software That Predicts Crime Pass Constitutional Muster?*, NPR (July 26, 2013, 4:55 PM), http://www.npr.org/2013/07/26/205835674/can-software-that-predicts-crime-pass-constitutional-muster, *archived at* http://perma.cc/C2GW-86LG (discussing predictive policing in Seattle).

[220] *See* JIE XU ET AL., RUTGERS CENT. ON PUB. SEC., CRIME GENERATORS FOR SHOOTINGS IN URBAN AREAS: A TEST USING CONDITIONAL LOCATIONAL INTERDEPENDENCE AS AN EXTENSION OF RISK TERRAIN MODELING 2 (2010) (reporting that shootings are concentrated around certain terrain features); Rubin, *supra* note 24 ("For patrol officers on the streets, mapping software on in-car computers and hand-held devices would show continuous updates on the probability of various crimes occurring in the vicinity, along with the addresses and background information about paroled ex-convicts living in the area.").

[221] *See* Ferguson, *Predictive Policing*, *supra* note 29, at 265-69.

[222] *See generally* Ferguson, *Crime Mapping*, *supra* note 25, at 184-90 (providing an overview of crime mapping techniques).

[223] *See generally* JAMES J. WILLIS ET AL., POLICE FOUND., COMPSTAT IN PRACTICE: AN IN-DEPTH ANALYSIS OF THREE CITIES 2-5 (2003) (providing an overview of COMPSTAT technology); James J. Willis et al., *Making Sense of COMPSTAT: A Theory-Based Analysis of Organizational Change in Three Police Departments*, 41 LAW & SOC'Y REV. 147, 148 (2007) ("COMPSTAT, a management and technological system, . . . [c]ombine[s] cutting-edge crime analysis and geographical information systems with state-of-the-art managements principles . . . .").

[224] Ferguson, *Predictive Policing*, *supra* note 29, at 265-69.

[225] *See* Anthony A. Braga et al., *The Relevance of Micro Places to Citywide Robbery Trends: A Longitudinal Analysis of Robbery Incidents at Street Corners and Block Faces in Boston*, 48 J. RES. CRIME & DELINQ. 7, 11 (2011) ("Studies of the spatial distribution of robbery in urban environments have also revealed that a small number of micro places generate a disproportionate number of robberies. Certain high-risk facilities, such as bars, convenience stores, and banks, at particular places also tend to experience a disproportionate amount of robbery."); Lisa Tompson & Michael Townsley, *(Looking) Back to the Future: Using Space–Time Patterns to Better Predict the Location of Street Crime*, 12 INT'L J. POLICE SCI. & MGMT. 23, 24 (2010) ("Research has repeatedly demonstrated that offenders prefer to return to a location associated with a high chance of success instead of choosing random targets.").

these place-based vulnerabilities allows analysts or algorithms to predict the next area of likely criminal activity.[226]

The collected crime data also holds other keys to understanding the actions of criminals.[227] Criminals adopt certain modi operandi and generally are creatures of habit.[228] In fact, one reason why certain crimes encourage almost "contagious" criminal activity in the surrounding areas is because the same criminals (or groups of criminals) are doing the acts.[229] Identifying these patterns may well provide clues as to who was involved in the crimes.[230] Sometimes these patterns, in conjunction with other factors, such as bus routes, escape routes, weather patterns, paydays, license plates, and special events, may also reveal a likely offender.[231]

At a deeper level of sophistication, with enough data, police will be able to predict criminal networks from patterns or connections.[232] Just as companies can identify you and your interests and associates from past activities, law enforcement might be able to target criminal networks using

[226] *See* Erica Goode, *Sending the Police Before There's a Crime*, N.Y. TIMES, Aug. 16, 2011, at A11 (reporting on predictive policing in Santa Cruz, California).

[227] This predictive focus on individuals has already been adopted in other areas of the criminal justice system, most notably in predicting recidivism and for pretrial release. Risk assessment mechanisms used in dozens of jurisdictions rely on predictive formulas to judge which offenders should be released and their likelihood of reoffending. Berk, *supra* note 128, at 1074.

[228] *Cf.* Cynthia Rudin, *Predictive Policing: Using Machine Learning to Detect Patterns of Crime*, WIRED (Aug. 22, 2013, 3:07 PM), http://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime, *archived at* http://perma.cc/84SQ-RCBG ("The algorithm tries to construct a modus operandi (M.O.) of the offender. The M.O. is a set of habits that the offender follows and is a type of behavior used to characterize a pattern. The M.O. for the burglaries included factors like means of entry (front door, back door, window), day of the week, characteristics of the property (apartment, single family house), and geographic proximity to other break-ins.").

[229] Kate J. Bowers & Shane D. Johnson, *Who Commits Near Repeats? A Test of the Boost Explanation*, W. CRIMINOLOGY REV., Nov. 2004, at 12, 22.

[230] *See* Vikas Grover et al., *Review of Current Crime Prediction Techniques*, *in* APPLICATIONS AND INNOVATIONS IN INTELLIGENT SYSTEMS XIV 233, 233 (Richard Ellis et al. eds, 2007) ("Data is not just a record of crimes, it also contains valuable information that could be used to link crime scenes based on the modus operandi (MO) of the offender(s), suggest which offenders may be responsible for the crime and also identify those offenders who work in teams (offender networks) etc."); *cf.* Usama Fayyad et al., *From Data Mining to Knowledge Discovery in Databases*, AI MAG., Fall 1996, at 37, 39 ("Historically, the notion of finding useful patterns in data has been given a variety of names, including data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing. The term *data mining* has mostly been used by statisticians, data analysts, and the management information systems (MIS) communities.").

[231] *See* Rudin, *supra* note 228.

[232] *See* Steve Lohr, *How Privacy Vanishes Online*, N.Y. TIMES, Mar. 17, 2010, at A1 (discussing the "power of computers to identify people from social patterns").

similar pattern recognition technologies.[233] For example, police concerned with the manufacture of methamphetamine could glean valuable information from commercial sales data in a jurisdiction. Big data analysts might track all of the purchases of individuals who bought several of the component parts required to make the drug: lye, iodine, ephedrine (Sudafed), Drano, brake fluid, and lighter fluid.[234] Each of these products has a lawful use, but identifying the individuals who bought all of these products would be a valuable clue in determining who might also be making methamphetamine. Patterns of anonymous sales data alone might demonstrate the levels of meth manufacture taking place; identifying the actual identities of repeat purchasers would benefit investigators even more. Though merely a more sophisticated form of criminal profiling,[235] this possibility has drawn the interest of major players, including the FBI,[236] who see the potential of big data pattern matching.[237]

[233] *See, e.g.*, Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, at A1 (describing "a ticking privacy time bomb" where search engine data can reveal user identity); Ryan Singel, *Netflix Spilled Your* Brokeback Mountain *Secret, Lawsuit Claims*, WIRED (Dec. 17, 2009), http://www.wired.com/threatlevel/2009/12/netflix-privacy-lawsuit, *archived at* http://perma.cc/GU22-4LWD (suggesting that Netflix users can be identified based on their viewing history and movie ratings).

[234] *See* Jon Bardin, *Kentucky Study Links Pseudophedrine* [sic] *Sales, Meth Busts*, L.A. TIMES (Oct. 16, 2012), http://articles.latimes.com/2012/oct/16/news/la-heb-kentucky-counties-pseudophedrine-meth-busts-20121016, *archived at* http://perma.cc/EG2P-EZHN ("Using that data, researchers were able to determine how much of the drug was sold in each Kentucky county and compare it with the number of meth busts in local police logs. . . . The researchers found a significant association between pseudophedrine [sic] sales and meth busts: In any given county, an increase in pseudophedrine [sic] sales of thirteen grams per 100 people translated to an additional meth lab busted. The results suggest that the computer databases could actually be used to predict where drug busts are most likely to take place.").

[235] *See* Steinbock, *supra* note 18, at 13 ("Data mining differs from data matching in that it is concerned with *patterns* of characteristics and behavior and is often used for making predictive judgments. . . . Data mining is also called 'knowledge discovery,' 'pattern-matching,' and 'dataveillance.'" (footnotes omitted)); *id.* at 30 ("[A]lthough predictive profiling is not inconsistent with the Fourth Amendment, the factors used must indicate to the investigating officers (and, later, the reviewing court) the requisite degree of suspicion. Nothing suggests that these actors should defer to a computer algorithm for projecting that level of suspicion, but nothing rules out that possibility either." (footnote omitted)).

[236] *Building Safer Communities: The Importance of Effective Federal–Local Collaboration in Law Enforcement, Hearing Before the S. Comm. on the Judiciary*, 112th Cong. 35 (2011) (statement of Richard A. McFeely, Special Agent in Charge, Balt. Field Office, FBI) ("Since September 11, 2001, the FBI has shifted from a traditional crime-fighting agency into an intelligence-led, threat-driven organization, guided by clear operational strategies. Today's FBI is focused on predicting and preventing the threats we face while at the same time engaging with the communities we serve. This shift has led to a greater reliance on technology, collaboration, and information sharing.").

[237] Erin Murphy, *Databases, Doctrine & Constitutional Criminal Procedure*, 37 FORDHAM URB. L.J. 803, 830 (2010) ("But the use of databases to generate suspects represents a new kind of investigation altogether—whether based on particular information (e.g., 'who called this number')

Several jurisdictions have even compiled "bad guy lists" of individuals they predict will commit crimes in the future or are involved in ongoing criminal activity but have not yet been caught. As Rodney Moore, the Chief of Police of Charlotte–Mecklenburg, North Carolina, stated, "We could name our top 300 offenders. So we will focus on those individuals, the persons responsible for the criminal activity, regardless of who they are or where they live. . . . We're not just looking for crime. We're looking for people."[238] These unofficial lists are not based on ongoing observed actions, but instead derive from a suspect's links to known criminal actors or past alleged actions.

D. *Unprotected Data*

Big data remains largely under-regulated. This Section reviews the constitutional, statutory, and commercial restrictions imposed on the collection and use of information underlying big data.

As a constitutional matter, few limits exist on accessing and collecting personal data. The controlling Fourth Amendment standard, derived from *Katz v. United States*, asks whether an individual has an expectation of privacy that society would consider objectively reasonable.[239] This expectation of privacy test has little application to the information police collect about individuals who enter the criminal justice system (including convictions, arrests, or biographical information provided pursuant to the criminal process). It also has little application to information individuals knowingly expose to the public, as the Supreme Court has reasoned that this information does not deserve Fourth Amendment protection.[240] In addition, information given to private individuals who later turn it over to law enforcement is not protected under the theory that the risk of disclosure was assumed by revealing the information to another person.[241] Similarly, data

or upon predefined algorithms (e.g., 'who has traveled to these three countries and bought these two items within a one month period').").

[238] Mitchell, *supra* note 28.

[239] *See* Katz v. United States, 389 U.S. 347, 357-59 (1967) (stating that searches without judicial approval are per se unreasonable—"subject only to a few specifically established and well-delineated exceptions"—and that people are "entitled to know that [they] will remain free from unreasonable searches and seizures").

[240] *See id*. at 351 ("What a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection."); *see also* California v. Ciraolo, 476 U.S. 207, 213 (1986) ("The Fourth Amendment protection of the home has never been extended to require law enforcement officers to shield their eyes when passing by a home on public thoroughfares.").

[241] *See* United States v. Jacobsen, 466 U.S. 109, 114-15 (1984) ("[T]he fact that agents of the private carrier independently opened the package and made an examination that might have been impermissible for a government agent cannot render otherwise reasonable official conduct

given to commercial third parties, including banking records,[242] telephone call lists,[243] cell phone locations,[244] or Internet search or subscriber information[245]

unreasonable. The reasonableness of an official invasion of the citizen's privacy must be appraised on the basis of the facts as they existed at the time that invasion occurred."); Coolidge v. New Hampshire, 403 U.S. 443, 489 (1971) (reasoning that when the suspect's wife produced evidence for the police, "it was not incumbent on the police to stop her or avert their eyes"); Hoffa v. United States, 385 U.S. 293, 302 (1966) ("Neither this Court nor any member of it has ever expressed the view that the Fourth Amendment protects a wrongdoer's misplaced belief that a person to whom he voluntarily confides his wrongdoing will not reveal it."); Burdeau v. McDowell, 256 U.S. 465, 475-76 (1921) ("[The Fourth Amendment's] origin and history clearly show that it was intended as a restraint upon the activities of sovereign authority, and was not intended to be a limitation upon other than governmental agencies . . . ."); Richard A. Epstein, *Privacy and the Third Hand: Lessons from the Common Law of Reasonable Expectations*, 24 BERKELEY TECH. L.J. 1199, 1222 (2009) ("[T]he law is entitled to the evidence of every person, and it is hard to think of a criminal system that could survive a new-found ability of every person to bind the state by contracting out of the third-party rules.").

[242] United States v. Miller, 425 U.S. 435, 443 (1976) ("The depositor takes the risk, in revealing his affairs to another, that the information will be conveyed by that person to the Government."). *Compare* Commonwealth v. Duncan, 817 A.2d 455, 463 (Pa. 2003) (reasoning that with respect to bank record disclosures, "[a] person's name and address do not, by themselves, reveal anything concerning his personal affairs, opinions, habits or associations." (internal quotation marks omitted)), *with* State v. McAllister, 875 A.2d 866, 874 (N.J. 2005) ("[B]ank records are simply a collection of numbers, symbols, dates, and tables. . . . However, when compiled and indexed, individually trivial transactions take on a far greater significance. . . . Indeed, the totality of bank records provides a virtual current biography." (internal quotation marks omitted)).

[243] Smith v. Maryland, 442 U.S. 735, 744 (1979) ("When he used his phone, petitioner voluntarily conveyed numerical information to the telephone company and 'exposed' that information to its equipment in the ordinary course of business. In so doing, petitioner assumed the risk that the company would reveal to police the numbers he dialed."); *see also* United States v. Christie, 624 F.3d 558, 574 (3d Cir. 2010) ("[N]o reasonable expectation of privacy exists in an IP address, because that information is also conveyed to . . . third parties, including ISPs."); United States v. Bynum, 604 F.3d 161, 164 (4th Cir. 2010) (holding that there is no objectively reasonable expectation of privacy in subscriber information given to an Internet service provider); United States v. Perrine, 518 F.3d 1196, 1204-05 (10th Cir. 2008) ("Every federal court to address [the] issue has held that subscriber information provided to an internet provider is not protected by the Fourth Amendment's privacy expectation."); United States v. Forrester, 512 F.3d 500, 510 (9th Cir. 2008) ("[E]-mail and Internet users have no expectation of privacy in the to/from addresses of their messages or the IP addresses of the websites they visit because they should know that this information is provided to and used by Internet service providers for the specific purpose of directing the routing of information.").

[244] *See In re* Historical Cell Site Data, 724 F.3d 600, 615 (5th Cir. 2013) (holding that individuals' historical cell location information stored by third-party cell providers is not protected by the Fourth Amendment). *See generally* Susan Freiwald, *Cell Phone Location Data and the Fourth Amendment: A Question of Law, Not Fact*, 70 MD. L. REV. 681, 702-08 (2011) (providing background information on government requests for location data).

[245] *See, e.g.*, United States v. D'Andrea, 497 F. Supp. 2d 117, 120 (D. Mass. 2007) ("The *Smith* line of cases has led federal courts to uniformly conclude that internet users have no reasonable expectation of privacy in their subscriber information, the length of their stored files, and other noncontent data to which service providers must have access.").

have not been protected by the third-party doctrine.[246] Some scholars have critiqued this policy, and Justice Sotomayor has expressed some inclination to reconsider the third-party doctrine.[247] By and large, however, Fourth Amendment protection is currently unavailable for this type of information.

Unlike these constitutionally unprotected categories of information, there exists a patchwork of statutes that limit the disclosure of health information, financial information, and some online communication.[248] To be clear, these statutes cover direct access to third-party information. For example, the Health Insurance Portability and Accountability Act of 1996[249] (HIPAA) protects access to medical records, although it allows law enforcement to access the records through an administrative, trial, or grand jury subpoena.[250] Likewise, laws such as the Gramm–Leach–Bliley Act,[251] the Bank Secrecy Act,[252] the Right to Financial Privacy Act of 1978,[253] and the Fair Credit Reporting Act,[254] provide some measure of protection from unauthorized access to financial records, although these protections can be surmounted by a subpoena or court order.

Similarly, the content of electronic communications is statutorily protected by the Electronic Communications Privacy Act of 1986[255] (ECPA)

[246] *See, e.g.*, Henderson, *Beyond the (Current) Fourth Amendment*, *supra* note 37, at 1015 ("Where the third party itself initiates the transfer, the 'private search' doctrine is controlling, in that the Fourth Amendment and its state analogues only restrict government conduct."); Henderson, *Fifty States*, *supra* note 37, at 395-96 (offering a fifty-state survey of states' positions on the federal third-party doctrine); Kerr, *supra* note 37, at 563 ("By disclosing to a third party, the subject gives up all of his Fourth Amendment rights in the information revealed."); Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801, 858 (2004) ("Because the Fourth Amendment reflects a clear commitment of the Framers to protect privacy, judges should identify the values of privacy in new technologies and translate them in to new Fourth Amendment rules." (footnote omitted)).

[247] United States v. Jones, 132 S. Ct. 945, 957 (Sotomayor, J., concurring) ("More fundamentally, it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties.").

[248] Murphy, *supra* note 38, at 503 (discussing the federal statutory limits on data disclosure—and corresponding exemptions for law enforcement).

[249] Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 26, 29, and 42 U.S.C. (2012)).

[250] *See* 45 C.F.R. § 164.512(f)(1)–(2) (2013).

[251] 15 U.S.C. § 6802 (2012).

[252] 12 U.S.C. §§ 1951–59 (2012).

[253] *Id.* at §§ 3401–22.

[254] 15 U.S.C. § 1681 (2012).

[255] Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended in sections of 18 U.S.C. (2012)); *but see* Communications Assistance for Law Enforcement Act, Pub. L. No. 103-414, 108 Stat. 4279 (1994) (requiring telecommunications carriers to maintain systems compatible with certain types of surveillance techniques).

and the Stored Communications Act,[256] but the protection lapses quickly.[257] Finally, telephone records are subject to protection through the Telephone Records and Privacy Protection Act of 2006,[258] but they too can be accessed by police if the evidence is relevant based on "specific and articulable" facts.[259]

In addition to constitutional and statutory protections, certain consumer guidelines established by companies promise to keep information private.[260] Yet most major commercial entities—including Internet search companies, online retailers, and social media platforms—collect data to monetize it.[261] In fact, many businesses, including big-name companies like Google, Microsoft, Yahoo!, and Facebook, are financially successful, in part, because of their ability to sell targeted advertising using user data.[262] These economic incentives, combined with a willingness to assist law enforcement as good corporate citizens, means that most third-party information is not well-protected from government access.

III. BIG DATA AND REASONABLE SUSPICION ON THE STREETS

What happens when a doctrine built on small data becomes overwhelmed by big data? What happens when previously unknown suspects can become known with a few quick search queries? Police and courts will soon confront this new reality as officers come to use existing facial recognition or biometric technology and networked databases to obtain individualized and particularized information about a suspect. Courts will confront additional questions as these technologies become more sophisticated, mobile, and reliant on predictive analytics.

[256] 18 U.S.C. §§ 2701–2712 (2012).

[257] *Compare* 18 U.S.C. §§ 2511, 2516, 2518 (2012) (describing the heightened requirements for obtaining real time communications), *with id.* § 2703(a) (setting out the lower standards for obtaining a court order for stored communications).

[258] 18 U.S.C. § 1039 (2012).

[259] *Id.* § 2703(c)–(d).

[260] *See, e.g., Microsoft.com Privacy Statement*, MICROSOFT, http://privacy.microsoft.com/en-us/default.mspx (last updated Aug. 2013), *archived at* http://perma.cc/F96M-8FUH; *Privacy Policy*, GOOGLE, http://www.google.com/privacy (last modified Mar. 31, 2014), *archived at* http://perma.cc/5FL4-NEHK;

[261] *See supra* Section II.B.2.

[262] *See, e.g.*, Rupert Neate & Rowena Mason, *Networking Site Cashes in on Friends*, TELEGRAPH (Jan. 31, 2009), http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/4413483/Networking-site-cashes-in-on-friends.html, *archived at* http://perma.cc/CBF6-R5N9 (reporting Facebook's move to monetize its collection of personal user information by allowing advertisers to target Facebook users selectively).

This Part studies this intersection of technology and doctrine through three different lenses—observation, investigation, and prediction—mirroring the most common types of police work. Police officers regularly observe ongoing criminal activity, investigate past criminal activity, and predict future criminal activity. The impact of "big data suspicion" will be different depending on the type of police activity at issue.

A. *Observation of Ongoing or Imminent Crimes*

Consider a modern day *Terry v. Ohio* situation. Detective McFadden is patrolling the street. He observes John Terry and, using facial recognition technology, identifies him and begins to investigate using big data. Detective McFadden learns through a database search that Terry has a prior criminal record, including a couple of convictions and a number of arrests.[263] McFadden learns, through pattern–matching links, that Terry is an associate (a "hanger on") of a notorious, violent local gangster—Billy Cox—who had been charged with several murders.[264] McFadden also learns that Terry has a substance abuse problem and is addicted to drugs.[265] These factors—all true, but unknown to the real Detective McFadden—are individualized and particularized to Terry. Alone, they may not constitute reasonable suspicion that Terry is committing or about to commit a particular crime. But in conjunction with Terry's observed actions of pacing outside a store with two associates, the information makes the reasonable suspicion finding easier and, likely, more reliable.

In observation cases, by using mobile facial recognition to identify the suspect, the officer now can turn any unknown suspect into a known suspect and can search for information that might justify reasonable suspicion. This change allows the officer to review traditional data sources known to law enforcement, including prior criminal history, arrests, addresses, gang associations, known associates, and even concealed weapons permits. Perhaps this individual is on a local "most wanted" list or a watch list as someone who has already been identified as being trouble in the neighborhood.[266] Perhaps

[263] Louis Stokes, *Representing John W. Terry*, 72 ST. JOHN'S L. REV. 727, 728-29 (1998) (discussing the facts of the *Terry* case); *see also* Terry v. Ohio *30 Years Later*, *supra* note 48, app.B at 1523 (reporting the sentencing judge as describing Terry as "a man who has from December 30, 1948, to the present time, be[en] consistently involved in difficulties with the law").

[264] Stokes, *supra* note 263, at 728-29.

[265] *Id.* at 727.

[266] *See* Slobogin, *Government Data Mining*, *supra* note 138, at 322 ("Match-driven data mining programs are designed to determine whether a particular individual has already been identified as a 'person of interest.' In other words, the goal here is not to find out more about a suspect, but rather to determine whether a particular person is a known suspect." (emphasis omitted)).

his height, weight, race, hairstyle, facial hair, or other distinguishing marks match a robbery suspect. This traditional law enforcement information might also now include data from automatic license plate readers, digitally archived surveillance video, and intelligence reports created and maintained by police. Even this limited information may be—as a constitutional matter—enough for an officer to stop the suspect. If, for example, the suspect had an extensive history of commercial robberies, or if license plate data connected him to prior robberies in the area, this information might well constitute reasonable suspicion that the suspect was going to commit a robbery.

Additional big data innovations may also assist the police. For example, the New York Police Department (NYPD) has unveiled the Domain Awareness System (DAS) in partnership with Microsoft.[267] This technology allows an officer to observe, through video surveillance or automated license plate readers, the location of a suspect prior to the initial observation:

> DAS is capable of rapidly blending and analyzing realtime data gathered from roughly 3,000 civic closed-circuit cameras, 911 call recordings, and license plate readers . . . as well as historical crime reports. Now the NYPD can do things like track a vehicle and instantly determine nearly everywhere it's been for the past few days or weeks; instantly access a suspect's arrest record, and all the 911 calls related to a particular crime; [and] map criminal history to geospatially and chronologically reveal crime patterns . . . .[268]

Thus, the officer could determine whether the suspect had just arrived with a getaway driver, had been casing the store, or had merely been doing non-criminal errands all morning.[269] These patterns may well affect whether an officer has reasonable suspicion that a suspect is about to commit a crime.

[267] *See* 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), *available at* http://news.microsoft.com/2012/08/08/new-york-city-police-department-and-microsoft-partner-to-bring-real-time-crime-prevention-and-counter terrorism-technology-solution-to-global-law-enforcement-agencies.

[268] Douglas Page, *Crime Fighting's Next Big Deal*, OFFICER.COM (Sept. 4, 2012), http://www.officer.com/article/10773317/crime-fightings-next-big-deal, *archived at* http://perma.cc/YTF5-A2UC; *see also* Michael Endler, *NYPD, Microsoft Push Big Data Policing Into Spotlight*, INFO. WK. (Aug. 20, 2012), http://www.informationweek.com/security/privacy/nypd-microsoft-push-big-data-policing-in/240005838, *archived at* http://perma.cc/DK97-7HMD (describing how DAS could lead to earlier apprehension of criminals).

[269] Somini Sengupta, *Privacy Fears as Surveillance Grows in Cities*, N.Y. TIMES, Oct. 13, 2013, at A1 (pointing out that big data-driven policing in Oakland, California, could help separate innocent actions from criminal activity).

For a second level of inquiry, imagine the police officer uses networked databases owned by third parties to discover personal information about a suspect. This data might include credit information, financial records, credit card activity, employment, past addresses and telephone numbers, names and addresses of family members, neighbors' addresses and telephone numbers, business associates, make, model, and color of registered vehicles, social security numbers, dates of birth, bankruptcies, liens and judgments, and GPS locational data. While access to some of these data would usually require particular legal authorization, law enforcement can circumvent statutes restricting direct access by instead using "fourth-party" commercial aggregators.[270] Such personalized information will allow an officer to develop a more individualized picture of a suspect. While generally unemployment, credit card debt, and bankruptcy are not indicia of criminal activity, when viewed in conjunction with suspicious action in front of an expensive jewelry store, however, a personal financial crisis might be relevant to the totality of circumstances. Further, accurate GPS data tying the suspect to a prior robbery or to a pawnshop might lead to reasonable suspicion. Even the otherwise innocent purchase of a wool cap or ski mask at Walmart might tip a seasonal purchase into reasonable suspicion.

Finally, imagine if law enforcement could access the suspect's social media data.[271] Search queries, Facebook and Twitter posts, YouTube videos, emails, texts, and similar communications are all available to third-party providers—if not publically available. While personal content is usually statutorily (or commercially) protected, it is generally not constitutionally protected. This mosaic of personal information might well provide individualized facts necessary to make the police officer's suspicion reasonable.[272] For example, a suspect's admission of financial difficulties or photograph displaying the fruits of the crime through social media could appropriately be added to the totality of circumstances.

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. A

[270] *See* Simmons, *supra* note 173, at 990-92 (describing commercial data acquisitions by the government).

[271] *See* Joseph Goldstein & J. David Goodman, *Seeking Clues to Gangs and Crimes, Detectives Follow Internet Rap Videos*, N.Y. TIMES, Jan. 8, 2014, at A20 ("Directed by prosecutors to build evidence that individual shootings are part of larger criminal conspiracies, officers are listening to local rappers for a better sense of the hierarchy of the streets. 'You really have to listen to the songs because they're talking about ongoing violence.'").

[272] *Cf. id.* (highlighting police use of social media to gain insight into criminal conspiracies).

trip to a pawnshop could indicate a person is selling stolen goods—or is merely poor enough to have to sell belongings at a steep discount. A photograph of jewelry could be an admission of theft or could simply be a photograph of jewelry. Yet in a criminal investigation, the inferences of suspicion are easy to develop and, against a low legal threshold, easy to meet.

Of course, suspicious facts must be connected with a suspected crime. It would not be relevant if the searches revealed a pattern of domestic violence crimes, unrelated to robbery. It would also not be relevant if the information was not directly connected to the suspect. Being a friend of a friend of a known robber is a fact, but not one that should influence the constitutional calculus. But, as long as the data are connected to both the suspected criminal activity and the suspected criminal, it would likely be persuasive in evaluating reasonable suspicion in observation cases.

B. *Investigation of Completed Crimes*

Many crimes occur without direct police observation, and police must investigate the crime to identify the perpetrator. Reasonable suspicion is still relevant in investigating past crimes (assuming the information available does not rise to the higher level of probable cause).[273] In *Hensley*, the Supreme Court set out the standard for investigating past crimes based on reasonable suspicion:

> The precise limits on investigatory stops to investigate past criminal activity are more difficult to define. The proper way to identify the limits is to apply the same test already used to identify the proper bounds of intrusions that further investigations of imminent or ongoing crimes. That test, which is grounded in the standard of reasonableness embodied in the Fourth Amendment, balances the nature and quality of the intrusion on personal security against the importance of the governmental interests alleged to justify the intrusion. When this balancing test is applied to stops to investigate past crimes, we think that probable cause to arrest need not always be required.[274]

[273] Illinois v. Gates, 462 U.S. 213, 241 (1983) ("[P]robable cause deals with probabilities." (internal quotation marks omitted)). The impact of big data on probable cause is a separate subject beyond the scope of this Article.

[274] United States v. Hensley, 469 U.S. 221, 228 (1985) (citations omitted); *see also id.* at 227 ("This is the first case we have addressed in which police stopped a person because they suspected he was involved in a completed crime. In our previous decisions involving investigatory stops on less than probable cause, police stopped or seized a person because they suspected he was about to commit a crime, or was committing a crime at the moment of the stop." (citation omitted)); *id.* ("We do not agree . . . that our prior opinions contemplate an inflexible rule that precludes police

While acknowledging that courts might balance these interests differently when investigating a past, completed crime—as opposed to an ongoing crime[275]—the Supreme Court still held that "the ability to briefly stop that person, ask questions, or check identification in the absence of probable cause promotes the strong government interest in solving crimes and bringing offenders to justice."[276] By adopting a reasonable suspicion test for investigation of past crimes, the Court gave police the flexibility to stop suspects based on this lower threshold of suspicion.[277]

As in observation cases, the primary use of big data would be to identify unknown perpetrators for arrest and prosecution. As one security expert explained, "[i]magine the ability to instantly take a security camera photograph from a bank robbery and match it using a facial recognition algorithm to a photograph in an out-of-state motor vehicle database, and then to link that person's name to a mobile phone from a private-sector marketing database."[278] Already, police have relied on similar linkages of networked information in more run-of-the-mill cases.[279] With new search technology, disparate pieces of data are compiled to link, match, and identify a suspect through pattern matching techniques. This can be done not only with a name, address, or license plate, but also with a particular modus operandi.[280]

from stopping persons they suspect of past criminal activity unless they have probable cause for arrest. To the extent previous opinions have addressed the issue at all, they have suggested that some investigative stops based on a reasonable suspicion of past criminal activity could withstand Fourth Amendment scrutiny.").

[275] *See id.* at 228-29 ("The factors in the balance may be somewhat different when a stop to investigate past criminal activity is involved rather than a stop to investigate ongoing criminal conduct.").

[276] *Id.* at 229.

[277] United States v. Place, 462 U.S. 696, 702 (1983) (allowing stops "when the officer has reasonable, articulable suspicion that the person *has been*, is, or is about to be engaged in criminal activity." (emphasis added)); Florida v. Royer, 460 U.S. 491, 498 (1983) (allowing certain seizures "if there is articulable suspicion that a person *has committed* or is about to commit a crime" (emphasis added)); United States v. Cortez, 449 U.S. 411, 417 n.2 (1981) ("Of course, an officer may stop and question a person if there are reasonable grounds to believe that person is wanted for past criminal conduct.").

[278] Page, *supra* note 268.

[279] *See* Mark Ward, *Crime Fighting with Big Data Weapons*, BBC (Mar. 18, 2014, 2:35 AM), http://www.bbc.com/news/business-26520013, *archived at* http://perma.cc/4ETS-GKDF; *see also* Neal Ungerleider, *This Small City's Police Department Builds an App, Nabs Big Data to Find and Fight Bad Guys*, FAST COMPANY (Mar. 26, 2014, 9:00 AM), http://www.fastcompany.com/3027641/this-small-citys-police-department-builds-an-app-nabs-big-data-to-find-and-fight-bad-guys, *archived at* http://perma.cc/7Z7H-5TNP.

[280] Taipale, *supra* note 184, at 21 ("The popular view of investigation in law enforcement is that there must first be a specific crime and that law enforcement then follows particularized clues or suspicions after the fact. In reality, investigators often look for criminal patterns or hypothetical suspects in order to anticipate future crime. For example, investigators may use pattern recogni-

This information, specific to a person and particularized to a crime, meets both requirements needed to establish reasonable suspicion.

The value of big data to reasonable suspicion investigations is probably greater than its value to observation cases, because police have time to surmount the "legal process" requirements necessary to obtain third-party information.[281] With an official request, a court order, or a subpoena (let alone a warrant or grand jury subpoena), law enforcement officers can obtain most third-party data if doing so in furtherance of a criminal investigation.[282]

Software can isolate patterns and identify suspects through existing public and private data in novel ways. One fascinating example of big data sleuthing arose out of the investigation of a major Swedish armed robbery of millions of dollars.[283] Police assumed that, to disguise their plot, the thieves must have used prepaid disposable phones. Data analysts then searched through the list of all prepaid disposable phones in the area looking for "a set of phones that stayed within their own miniature network."[284] Police analysts found a single set of phones that only communicated with each other, did so only for a few weeks leading up to the heist, and then went silent after the robbery. Identifying this network allowed police to solve the case. Police traced the phones to specific cell tower locations corresponding with the robbers' locations before, during, and after the robbery.[285] In fact, once police knew the numbers, they could track location-by-location exactly where the robbers had been. When police identified one person who had purchased the phones, they were able to determine how the crime occurred and the location of the thieves at all times.[286]

Major police departments, as well as the FBI, have adopted this type of pattern matching investigation technique.[287] In child abduction cases,

tion strategies to develop modus operandi ('MO') or behavioral profiles, which in turn may lead either to specific suspects (profiling as identifying pattern) or to crime prevention strategies (profiling as predictor of future crime, resulting, for example, in stakeouts of particular places, likely victims, or potential perpetrators).").

[281] *See supra* Section II.D (discussing the statutory requirements of court orders for some private information).

[282] *See supra* Section II.D (noting the ease with which law enforcement may access records that are protected by statute).

[283] EVAN RATLIFF, LIFTED ch. 5–9 (Kindle Singles ed. 2011), *available at* https://www.atavist.com/stories/lifted/ (describing the investigation of the heist).

[284] *Id.* at ch. 12.

[285] *Id.*

[286] *Id.* at ch. 13.

[287] *See* Josh Richman & Angela Woodall, *Around the Bay Area, You're Being Watched*, CONTRA COSTA TIMES (June 30, 2013, 1:29 AM), http://www.contracostatimes.com/News/ci_23569173/Around-the-Bay-Area-youre-being, *archived at* http://perma.cc/V8UE-2TJR ("[I]t's not just the

"Amber alerts" have led to quick reviews of license plate reader databases. By searching the location of a car, police can determine the likely route of the suspect.[288] In gang cases, recordings of gunshots have helped map out areas of contested gang turf.[289]

Returning to the robbery example, imagine that a particular jewelry store was robbed by an unknown suspect. Police officers have a video still from the robbery that does not allow for a facial recognition match. The photo, however, clearly shows a neck tattoo, and officers obtain a partial description of the getaway car. Running a search for the tattoo against a database might narrow the list of suspects. Comparing the narrowed list with owners of a particular type of car might further limit the list of suspects. Looking at the remaining suspects' associates, movements, or even bank deposits, credit card expenditures, or social media comments might again tighten the search. The result is that big data can help identify the suspect with a few search queries. While these data might not be enough to get an arrest warrant, they would likely provide the reasonable suspicion needed to stop and investigate the suspect.[290]

C. *Predicting Crimes*

Unlike observation or investigation cases, reasonable suspicion based on prediction remains the stuff of science fiction. Police have begun to predict areas of heightened criminal activity,[291] and may predict likely troublemakers

National Security Agency secretly vacuuming up your personal data. Local police agencies are increasingly adopting Big Data technologies . . . ."); *cf.* Charles Piller & Eric Lichtblau, *FBI Plans to Fight Terror with High-Tech Arsenal*, L.A. TIMES, July 29, 2002, at A1 ("By Sept. 11, 2011, the FBI hopes to use artificial-intelligence software to predict acts of terrorism the way the telepathic 'precogs' in the movie 'Minority Report' foresee murders before they take place.").

[288] Lochner, *supra* note 199, at 225 ("[T]he Automated License Plate Recognition system, store[s] license plate numbers of the innocent and guilty so the database can be mined during Amber Alerts or for leads in cases.").

[289] *See* Christopher Benjamin, Note, *Shot Spotter and FaceIt: The Tools of Mass Monitoring*, UCLA J.L. & TECH., Spring 2002, art. 2, at 6 (describing a system by which automated phone calls help find the location of gunfire).

[290] *See* Cook, *supra* note 18 ("The Boston Police Department is rolling out a powerful new computer program built to find hidden connections among people and events almost instantly, allowing detectives to investigate murders, rapes, and other crimes far faster than they can today."); *see also* Yang Xiang et al., *Visualizing Criminal Relationships: Comparison of a Hyperbolic Tree and a Hierarchical List*, 41 DECISION SUPPORT SYS. 69, 75-77 (2005) (describing how a tool known as COPLINK Criminal Relationship Visualizer links co-occurring events and characteristics).

[291] Ferguson, *Predictive Policing*, *supra* note 29, at 312-13; Paul Bowers, *Predictive Policing Arrives in Charleston*, CHARLESTON CITY PAPER (June 27, 2012), http://www.charlestoncitypaper.com/charleston/predictive-policing-arrives-in-charleston/Content?oid=4101684, *archived at* http://perma.cc/JWL7-35TD (discussing the use of predictive analytics to reduce armed robberies in Charleston, South Carolina).

involved in criminal enterprise through an unofficial "bad guy list," but predictive analytics cannot yet tell police whom to stop for a crime not yet committed. To be clear, these are prediction-based stops where no crime has occurred and no crime is observed.

Yet big data invites provocative questions about whether such predictive tips should factor into the reasonable suspicion calculus. For example, if a drug distribution gang is run by a tight-knit family or neighborhood organization, such that the pattern for several years has been that when one family member is arrested, another cousin or brother takes their place, then why can we not predict who will be the next member of the gang?[292] If burglaries are contagious in part because the same gang of burglars commits similar crimes, and police identify one burglar, why should we not target a burglar's associates as likely suspects for future burglaries?[293] In these cases, police could show specific and articulable facts indicating that a particular person is likely to participate in ongoing criminal activity (e.g., drug dealing or burglaries).[294] Because the criminal enterprise is ongoing, the *Terry* standard might well apply, and police could try to stop and investigate would-be members of these criminal organizations if they were observed doing anything that might suggest drug dealing or burglary.

The questions get harder when no ongoing criminal enterprise exists, yet the same predictive logic holds. In Chicago, analysts have identified young people at greater risk of being involved in gun violence.[295] Researchers

[292] *See* STEVEN D. LEVITT & STEPHEN J. DUBNER, FREAKONOMICS: A ROGUE ECONOMIST EXPLORES THE HIDDEN SIDE OF EVERYTHING 89-114 (2005) (discussing the economics and social relationships of the drug trade in the famous chapter "Why Do Drug Dealers Still Live with Their Moms?").

[293] *See* Wim Bernasco, *Them Again?: Same-Offender Involvement in Repeat and Near Repeat Burglaries*, 5 EUR. J. CRIMINOLOGY 411, 423-25 (2008) ("[B]oth repeat burglaries and near repeat burglaries are much more likely to involve the same offender than are spatially or temporally unrelated burglaries."); Bowers & Johnson, *supra* note 229, at 13 (discussing how features of an offender's modus operandi, like spatial and temporal preferences, can be used to identify crimes carried out by a particular network of offenders).

[294] Domestic violence also presents a possible predictive environment for crime. *See* Joseph Goldstein, *Police Take on Family Violence to Avert Death*, N.Y. TIMES, July 25, 2013, at A1 ("[T]he officers assigned to the domestic violence unit make a total of 70,000 precautionary visits a year to the households with past episodes. Each precinct station house also maintains a 'high propensity' list of a dozen or so households that get special attention because they are believed to be most at risk of further violence.").

[295] *See* Andrew Papachristos, Tracey L. Meares & Jeffrey Fagan, *Attention Felons: Evaluating Project Safe Neighborhoods in Chicago* 4 J. EMPIRICAL LEGAL STUD. 223, 229-33 (describing Chicago's program to identify and address likely perpetrators and victims of gun violence); *see also* TRACEY MEARES, ANDREW V. PAPACHRISTOS & JEFFREY FAGAN, HOMICIDE AND GUN VIOLENCE IN CHICAGO: EVALUATION AND SUMMARY OF THE PROJECT SAFE NEIGHBORHOODS PROGRAM 1 (2009), *available at* http://www.psnchicago.org/PDFs/2009-PSN-Research-Brief_v2.pdf ("Data analysis

can predict their likelihood of being a victim or perpetrator of gun violence using big data metrics, including place of residence, social associations (e.g., past experience with victims of gun violence and gang connections), and age.[296] Assuming the accuracy of these data, could police target these individuals as part of a predictive stop strategy?[297] In fact, the Chicago Police Department appears to have adopted this predictive logic in its intervention program. As described by the *New York Times*,

> [i]n recent months, as many as 400 officers a day, working overtime, have been dispatched to just 20 small zones deemed the city's most dangerous. The police say they are tamping down retaliatory shootings between gang factions by using a comprehensive analysis of the city's tens of thousands of suspected gang members, the turf they claim and their rivalries. The police are also focusing on more than 400 people they have identified as having associations that make them the most likely to be involved in a murder, as a victim or an offender.[298]

immediately revealed that a very small number of neighborhoods in Chicago are responsible for most of the city's violence trends. The 'city's' crime problem is in fact geographically and socially concentrated in a few highly impoverished and socially isolated neighborhoods. Data also revealed that most victims (and offenders) of gun violence in Chicago tend to be young African American men who live in neighborhoods on the West or South sides of the city.").

[296] John Buntin, *Social Media Transforms the Way Chicago Fights Gang Violence*, GOVERNING, Oct. 2013, at 26, 28 ("Today, the Chicago Police Department is doing something similar with gangs. Using a tool academics call 'network analysis,' the CPD is mapping the relationships among Chicago's 14,000 most active gang members. It's also ranking how likely those people are to be involved in a homicide, either as victims or offenders. In the process, the CPD has discovered something striking: Cities don't so much have 'hot spots' as 'hot people.' That finding is transforming the way the police do business in Chicago and has significant implications for how other cities should be policed.").

[297] Michael Sierra-Arevalo, *How Targeted Deterrence Helps Police Reduce Gun Deaths*, SCHOLARS STRATEGY NETWORK (June 3, 2013, 1:11 PM), http://thesocietypages.org/ssn/2013/06/03/targeted-deterrence, *archived at* http://perma.cc/GZ65-U25X ("The perpetrators of gun violence are also concentrated in particular sectors of the population. In places like Boston, more than 50% of all murders and 70% of all shootings are committed by about one percent of youth aged 15 to 24."); *see also id.* ("Initiatives like The Boston Gun Project and Chicago's Project Safe Neighborhoods allow police to concentrate their efforts on gang-affiliated individuals with previous criminal records."). *See generally* OFFICE OF JUVENILE JUSTICE & DELINQUENCY PREVENTION, U.S. DEP'T OF JUSTICE, PROMISING STRATEGIES TO REDUCE GUN VIOLENCE 26-33 (1999), *available at* http://www.cops.usdoj.gov/html/cd_rom/solution_gang_crime/pubs/PromisingStrategiestoReduceGunViolence.pdf (discussing Boston's strategy to reduce gun violence by targeting specific groups and geographic areas).

[298] Monica Davey, *Chicago Tactics Put a Major Dent in Killing Trend*, N.Y. TIMES, June 11, 2013, at A1; *see also* Mark Guarnio, *Can Math Stop Murder?*, CHRISTIAN SCI. MONITOR (July 20, 2014), http://www.csmonitor.com/USA/2014/0720/Can-math-stop-murder-video, *archived at* http://perma.cc/G3TA-9SPT (discussing predictive policing techniques in Chicago including sending officers to the houses of suspected gang leaders).

Those four hundred individuals—part of a list of predicted offenders—were identified through big data techniques. Chicago police call it a "heat list."[299] Young men on the heat list are targets of predictive intervention-based strategies.

While a Fourth Amendment stop based solely on an individual's inclusion on this list, without more, might not be sufficiently particularized, big data tools exist to generate the necessary reasonable suspicion.[300] For example, imagine one of those four hundred individuals is a young man whom police wish to stop because they suspect that he is up to no good (and likely in possession of a gun). Plainly, an officer's suspicion that someone is "up to no good" does not constitute constitutionally sufficient justification for a stop. An officer sees the young man on the streets (but not engaged in any overt criminal activity). The officer identifies the young man as being on a list of individuals that predictive analytics suggested are at a heightened risk of involvement in gun violence. A quick NCIC database search reveals gang contacts, criminal associates, and prior arrests—including gun charges. Gang tattoos link the young man to local gangs. A license plate reader places the family car in the general vicinity of a gang shooting in the last month. His Facebook profile contains statements that police could interpret as directing violence at rival gang members.[301] Finally, predictive policing software has forecast the young man's location as the site of likely gun violence. If the police officer stops the young man (doing nothing overtly criminal) and finds a gun during a frisk, would a court really say there was not individualized and particularized suspicion that this individual was involved in gun and gang-related activity? Though the young man took no action to signify criminal activity, the data suggest that he was far more likely to be in possession of a gun than most people in Chicago.

How courts resolve these issues will determine the impact of big data on law enforcement. On one hand, judges might require some affirmative, imminent suspicious activity correlating with gun possession before upholding the stop, such as "furtive movements," a suspicious bulge, or unexplained

[299] Jeremy Gorner, *Chicago Police Use "Heat List" As Strategy to Prevent Violence*, CHI. TRIB. (Aug. 21, 2013), http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list, *archived at* http://perma.cc/8TJA-Y6KM.

[300] Presence on the list might also allow police to identify individuals for whom therapeutic intervention might be necessary.

[301] *See, e.g.*, JAAP BLOEM ET AL., SOGETI TREND LAB VINT, BIG SOCIAL: PREDICTING BEHAVIOR WITH BIG DATA 35 (2012), *available at* http://blog.vint.sogeti.com/wp-content/uploads/2012/10/Big-Social-Predicting-Behavior-with-Big-Data.pdf ("In the Netherlands, police officers go on duty with a smartphone in order to be able to pick up signals in the neighborhood from social media. In this way, they can show their faces before something serious happens in the schoolyard, for example.").

nervousness.[302] Without the requirement of some observable activity, the odds increase that predictive stops will target innocent people, criminalize by association, and negatively impact individuals based on little more than a hunch supported by non-criminal facts. On the other hand, many judges might find this totality of suspicions—even if focused on a particular suspect and not a crime—sufficient to justify an investigatory stop. Reasonable suspicion is a low threshold. Thus, in practice, aggregated reasonable suspicion would likely justify a stop in many courtrooms. As Part IV explains, this shift has significant implications for the Fourth Amendment.

D. *Big Data Suspicion*

Big data's ability to generate information about an identified suspect reveals the inherent vulnerability in the reasonable suspicion standard. Indeed, along the continuum of suspicion, more data makes it easier to satisfy the standard for two primary reasons. First, under a totality-of-circumstances test, the more factors a court considers in the totality, the easier it is to articulate suspicion. Quantity can make up for quality.[303] Second, the information provided by big data is individualized and particularized, consistent with the *Terry* language.[304] To be clear, the data are individualized to the criminal, not the crime. As courts apply *Terry*, however, which arose in the unknown suspect context, the difference becomes blurred.

This latter point is important to emphasize. The language in the earlier reasonable suspicion cases speaks to a general suspicion of unspecified criminal actions, using terms like "criminal activity may be afoot,"[305] "involved in criminal activity,"[306] "legal wrongdoing,"[307] or "illegality."[308] The general language does not require discussion of a particular observed crime (e.g., drug distribution or gun possession), because the officer actually observed the illegal activity in question. The observed crime and the

[302] *See, e.g.*, Jackson v. United States, 56 A.3d 1206, 1209-12 (D.C. 2013) (discussing the difficulty of interpreting furtive gestures and nervousness).

[303] Jane Bambauer, *Hassle*, 113 MICH. L. REV. (forthcoming 2014) (manuscript at 5), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2404088 (recognizing "courts' consistent preference for police narratives chock full of detail, even when each additional detail does not contribute much to the suspicion"); *see also id.* (manuscript at 42) ("When assessing an officer's decision to stop or search somebody, courts prefer a long lists [sic] of reasons. The more reasons the agent can recount, the better.").

[304] Terry v. Ohio, 392 U.S. 1, 21 (1968).

[305] *Id.* at 30.

[306] Brown v. Texas, 443 U.S. 47, 51 (1979).

[307] United States v. Arvizu, 534 U.S. 266, 273 (2002).

[308] Florida v. J.L., 529 U.S. 266, 272 (2000).

observed criminal were not separate things to analyze; no distinction was needed in the analysis. Thus, in a small data world, the traditional language describing suspicious behavior has no meaning outside of the observed activity.

In a big data world, this same generalized language becomes distorted. An officer may know information about a suspect, but the question becomes: how does that information relate to the observed actions? Knowing someone is a "drug dealer" does not mean that the individual is actively dealing drugs at the moment of observation. Courts analyzing big data suspicion should thus be careful to require a direct link between the past data about a suspect and the observed suspicion. With big data suspicion, it is important for the individualized and particularized information to relate to the particular action observed. If a police officer identifies a suspect and learns information about the suspect's arrests, convictions, or associations that has nothing to do with the observed actions (if the officer observed any actions at all), then the new information should be irrelevant to the reasonable suspicion calculus. Only when those particularized factors can be connected to observed actions that signify criminal activity should they affect the analysis.

Courts will soon be asked to address the impact of big data on reasonable suspicion. But before that time, policymakers will need to think through and evaluate whether this innovation is good or bad for police, individuals, and society.

IV. EVALUATING BIG DATA AND PREDICTIVE REASONABLE SUSPICION

While big data may expose the fragility of the reasonable suspicion doctrine, the technology has arrived, and its impact on Fourth Amendment cases is imminent. As such, it is necessary to evaluate the questions of law and policy that arise from the move to big data policing. This Part discusses positives and negatives of big data policing and provides suggestions on how to address the pending evolution of Fourth Amendment doctrine.

A. *The Positives of Big Data Suspicion*

As may be evident from early adoption and experimentation with predictive technologies, law enforcement officials see the potential of these tools to reduce crime. Big data suspicion, if used correctly, can improve accuracy and efficiency, and it will yield unexpected insights into the patterns of criminal activity.

1. Improved Accuracy

Reasonable suspicion based on big data primarily benefits law enforcement because of the increased accuracy it purports to offer.[309] More information, and more precise information, should make it more likely that police target actual criminals rather than innocent people. In a small data environment, police rely on proxies for information to the detriment of everyone. Class, race, age, choice of clothing, and gender all factor into police officers' discretionary decisions on the street.[310] Police perceive ambiguous actions as suspicious because of subtle cues or instincts. These judgments also unfortunately include explicit and implicit biases, policing traditions, and the frailties of human perception.[311] Replacing those generalized intuitions with

[309] *Cf.* Rachael King, *IBM Analytics Help Memphis Cops Get "Smart,"* BLOOMBERG BUSINESSWEEK (Dec. 05, 2011), http://www.businessweek.com/technology/ibm-analytics-help-memphis-cops-get-smart-12052011.html, *archived at* http://perma.cc/Q77C-WCXW (describing the technology used by law enforcement in Memphis, Tennessee, which has contributed to the lowest crime rates there in a quarter-century).

[310] *See, e.g.*, Shima Baradaran, *Race, Prediction, and Discretion*, 81 GEO. WASH. L. REV. 157, 200 (2013) (examining "whether police demonstrate racial bias" in deciding whether to make arrests); Katherine Y. Barnes, *Assessing the Counterfactual: The Efficacy of Drug Interdiction Absent Racial Profiling*, 54 DUKE L.J. 1089, 1113, 1132-35 (2005) (explaining study results in which a driver's race was found to be "the most salient factor" in deciding whether to search a vehicle); Angela J. Davis, *Race, Cops, and Traffic Stops*, 51 U. MIAMI L. REV. 425, 425 (1997) (describing the reluctance of two black men to draw additional attention to themselves while driving because their race and gender already "makes them more likely to be stopped and detained by the police"); David A. Harris, Essay, *"Driving While Black" and All Other Traffic Offenses: The Supreme Court and Pretextual Traffic Stops*, 87 J. CRIM. L. & CRIMINOLOGY 544, 546, 570 (1997) ("[P]retextual police stops of blacks are so common—frequent enough to earn the name "driving while black"—[that] many African-Americans regularly modify the most casual aspects of their driving behavior . . . and even their personal appearance . . . ."); Noel Leader, Panel Discussion at CUNY School of Law (Sept. 29, 2010), *in Suspect Fits Description: Responses to Racial Profiling in New York City*, 14 CUNY L. REV. 57, 65-67 (2010) (asserting that illegal stops based on racial profiling are breaches of officers' duty, though police often attempt to justify them by citing alternative explanations like the suspect's dress); Tracey Maclin, Terry v. Ohio*'s Fourth Amendment Legacy: Black Men and Police Discretion*, 72 ST. JOHN'S L. REV. 1271, 1279-87 (1998) (arguing that although methods in place in the 1960s to deter crime were facially race-neutral, the implementation of these strategies was largely determined by the race of the subject).

[311] *See* 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, 15-16 (2010) ("It is true that some people are, at times, reasonably good at making certain judgments based on first impressions. But they are also often quite bad at doing so. Moreover, first impressions can involve at least five major attributes, namely, the subject's emotions, personality, intelligence, mental states, and use of deception."); *id.* at 16 ("In addition, individuals' self-knowledge about the relative degree of accuracy of their ability to make judgments concerning each of the five major attributes upon first impression is also poor."); *see also* L. Song Richardson, *Police Efficiency and the Fourth Amendment*, 87 IND. L.J. 1143, 1147 (2012) ("Implicit social cognition research demonstrates that people have nonconscious reactions to others that can negatively influence their behaviors. These implicit biases begin when people categorize others both consciously and nonconsciously by

precise detail about actual people should result in a more accurate policing strategy. Humans are notoriously bad at making snap judgments, and while police officers make more snap judgments than most, they are not immune from the imperfections of human nature.[312]

For example, stories of racial profiling involving famous celebrities, wealthy professionals, and other citizens show how racial stereotypes can influence suspicion.[313] Yet in a big data world, a quick license plate scan or facial recognition check and a query of other databases (perhaps including professional licenses or even addresses), could help avoid the indignity of detention based solely on a police hunch.[314] Of course in many cases, information will not reveal that the individual is a celebrity, but even basic

race, gender, or a host of other socially relevant categories. Categorization triggers implicit stereotypes and attitudes." (footnotes omitted)).

[312] Eli B. Silverman, *With a Hunch and a Punch*, 4 J.L. ECON. & POL'Y 133, 140 (2007) ("Like other individuals within the same occupation, police vary in their ability to make intelligent, intuitive choices. Just as it varies among the general population, some police are better than others in detecting patterns from experience. Research and empirical observation amply demonstrates that there is a wide range in the ability of police officers to successfully deploy reasonable hunches in their work."); *see also* Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945, 947 (2006) (discussing the effects of mental processes outside of "conscious attentional focus" on decisionmaking); L. Song Richardson, *Cognitive Bias, Police Character, and the Fourth Amendment*, 44 ARIZ. ST. L.J. 267, 271 (2012) ("It is highly probable that fundamental attribution error affects police judgments of criminality. Officers on the beat often make quick decisions based upon limited evidence. The stressful nature of their jobs likely depletes their cognitive capacities, making correction for fundamental attribution error more difficult."); *id.* at 269 ("It is well established in the psychological literature that people tend to explain the behaviors of others by reference to their character (disposition) rather than to situational influences.").

[313] *See* CHARLES J. OGLETREE, JR., THE PRESUMPTION OF GUILT: THE ARREST OF HENRY LOUIS GATES JR. AND RACE, CLASS, AND CRIME IN AMERICA, 129-241 (2010) (telling the stories of one hundred influential African Americans who faced racial profiling or discrimination); David A. Harris, *The Stories, the Statistics, and the Law: Why "Driving While Black" Matters*, 84 MINN. L. REV. 265, 273-74 (1999) (describing measures taken by African Americans to avoid police harassment while driving); Sheri Lynn Johnson, *Race and the Decision To Detain a Suspect*, 93 YALE L.J. 214, 214 (1983) ("Thirty years ago police stopped Malcolm X because he was a black man in a white neighborhood. A revolution in civil rights later, police still view race as an important factor in the decision to detain a suspect." (footnote omitted)).

[314] *Compare* Albert W. Alschuler, *The Upside and Downside of Police Hunches and Expertise*, 4 J.L. ECON. & POL'Y 115, 118-19 (2007) (acknowledging that while hunches may be developed from real world experience, they are unreliable, shaped by racial stereotypes, burdensome to law enforcement, and unreviewable), *and* Harold Baer, Jr., *Got a Bad Feeling? Is That Enough? The Irrationality of Police Hunches*, 4 J.L. ECON. & POL'Y 91, 103 (2007) ("Until law enforcement agencies spend more time and money addressing the problems that arise from their culture, training and, in some locales, education, the hunch will remain problematical and occasionally unjust."), *with* Craig S. Lerner, *Judges Policing Hunches*, 4 J.L. ECON. & POL'Y 25, 25 (2007) ("[E]motions and intuitions are not obstacles to reason, but indispensable heuristic devices that allow people to process diffuse, complex information about their environment and make sense of the world.").

personal or employment data (or lack of criminal information) might provide police with a clue that a suspect is just an ordinary citizen not involved in criminal activity.

While vulnerable to abuse, predictive suspicion ultimately may make police stops more reliable. At its core, reasonable suspicion is a doctrine of predictive suspicion. The collected totality of circumstances must justify an officer's prediction that criminal activity is afoot.[315] Thus, having more information about an individual should result in more reliable predictions.[316] If police focus their efforts on people placed on "bad guy lists," it may protect individuals who are not on the lists. If police are forced to use big data to identify and link a suspect to a crime, they may also see patterns that suggest that the suspect was not involved in the crime. In this way, big data policing may be a measure more protective of individuals on the street.

The accuracy that big data provides not only increases the likelihood that police target the right suspects, but also, in turn, prevents the resulting physical, face-to-face interactions that generate tension.[317] Many police stops involve confirming or disproving suspicion.[318] Even if no arrest results, the unpleasant (and perhaps unnecessary) police–citizen contact breeds resentment and distrust.[319] Allowing police to confirm a person's lack

[315] *Cf.* Andrew E. Taslitz, *Fortune-Telling and the Fourth Amendment: Of Terrorism, Slippery Slopes, and Predicting the Future*, 58 RUTGERS L. REV. 195, 201 (2005) ("What is less often emphasized is that *Katz* faced the Justices with the question whether it is possible to authorize a search for non-existent evidence—evidence that may or may not come into being in the future.").

[316] *But cf.* Steinbock, *supra* note 18, at 38 ("The Fourth Amendment permits interferences with liberty and privacy based on predictions, often made by field officers, without notice to or consultation with the suspect.").

[317] *See* Frank Rudy Cooper, *"Who's the Man?": Masculinities Studies,* Terry *Stops, and Police Training*, 18 COLUM. J. GENDER & L. 671, 729-32 (2009) (criticizing police training programs for cultivating the culture of machismo and militarism that leads to police violence); James Forman, Jr., *Community Policing and Youth as Assets*, 95 J. CRIM. L. & CRIMINOLOGY 1, 14 (2004) (discussing police–citizen tension caused by "[b]elittling remarks, illegitimate orders, and cursing" by police during stops).

[318] *See* Minnesota v. Dickerson, 508 U.S. 366, 373 (1993) ("[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 . . . the officer may briefly stop the suspicious person and make reasonable inquiries aimed at confirming or dispelling his suspicions." (internal quotation marks omitted)).

[319] Mich. Dep't of State Police v. Sitz, 496 U.S. 444, 465 (1990) (Stevens, J., dissenting) ("[T]hose who have found—by reason of prejudice or misfortune—that encounters with the police may become adversarial or unpleasant without good cause will have grounds for worrying at any stop designed to elicit signs of suspicious behavior. Being stopped by the police is distressing even when it should not be terrifying, and what begins mildly may by happenstance turn severe."); 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) ("[I]t is precisely at this intersection of crime, race and, police stop and frisk practices that the underlying social and legal conflicts most often are manifested, and not infrequently in sharp and violent confrontations.").

of involvement in criminal activity through a database search, rather than a physical stop, avoids unnecessary conflict.

2. Exculpatory Facts

Suspicion is not a one-way street. Suspicion can be disproved. Suspicion can be alleviated. The advent of big data suspicion may require consideration of exculpatory factors that lessen suspicion. Just as big data enables a wealth of suspicious inferences, it also generates an equal number of potentially exculpatory facts. For example, if Detective McFadden learned that John Terry's wife worked near the downtown location of the observation, pacing outside a store might turn from "casing a robbery" to "waiting for a loved one."

The potentially exculpatory nature of big data is a strong positive argument for its use in policing. Presumably, if big data information exists about a suspect, police should be obligated to check before initiating a stop.[320] The totality of circumstances should not be understood as the totality of suspicious activities; it should include exculpatory information that reduces suspicion as well. This is an established part of the probable cause analysis,[321] and big data technology allows it to be included in the reasonable suspicion analysis. Thus, existing exculpatory information should be factored into the totality of circumstances and weighted just as heavily as suspicious factors.

Courts might even require police to factor in exculpatory information as a self-contained check on the regular discretionary powers granted to police. When big data is available, an officer who did not use it in an exculpatory manner might be deemed to have acted recklessly.[322] In the same way that courts may take a negative inference from an unrecorded confession in a jurisdiction where videotaping confessions is the norm, a failure to use the available search technology might be held against the officer.[323] In this way,

[320] *See, e.g.*, Joshua A.T. Fairfield & Erik Luna, *Digital Innocence*, 99 CORNELL L. REV. 981, 1031 (2014) (arguing that defendants have a right to government-created exculpatory big data).

[321] *See, e.g.*, United States v. Grubbs, 547 U.S. 90, 95 n.2 (2006) ("[P]robable cause may cease to exist after a warrant is issued. The police may learn, for instance, that contraband is no longer located at the place to be searched."); United States v. Watson, 423 U.S. 411, 432 n.5 (1976) (Powell, J., concurring) ("But in some cases the original grounds supporting the warrant could be disproved by subsequent investigation that at the same time turns up wholly new evidence supporting probable cause on a different theory.").

[322] *See, e.g.*, Andrew Guthrie Ferguson, *Constitutional Culpability: Questioning the New Exclusionary Rules*, 66 FLA. L. REV. 623, 648-52 (2014) (discussing recklessness in the context of Fourth Amendment violations).

[323] *Cf., e.g.*, Steven A. Drizin & Beth A. Colgan, *Let the Cameras Roll: Mandatory Videotaping of Interrogations Is the Solution to Illinois' Problem of False Confessions*, 32 LOY. U. CHI. L.J. 337, 385-88

using big data to determine reasonable suspicion might actually prevent certain stops that would have been allowed under a traditional, small data reasonable suspicion standard.

3. Accountability

Focusing on big data sources also provides the potential for increased documentation of stops. In general, police do not document their suspicions *before a stop*, nor does anyone do so on their behalf.[324] Data-driven suspicion, though, can be documented beforehand. Police officers could demonstrate the steps they took to investigate a suspect by producing records of which databases they accessed and which search queries they used. In this way, police would replace the ex post justification for a stop with an ex ante description of the steps taken to validate a hunch before conducting a stop. In simple terms, a police officer could show that she checked the NCIC database and ran a license plate check before explaining why this information corroborated her initial suspicion. This record has the potential not only to limit whom police stop, but also to make a judge's determination of an officer's reasonable suspicion significantly easier.

This documentation will also encourage the development of a culture that allows for auditing of data, standards for record collection, and perhaps even notice requirements for targeted suspects. For example, police administrators, as an internal monitoring strategy, might examine an officer's history of stops to see what factors influenced his decision to stop a suspect. Looking through the documented history of big data searches and comparing them with the justifications for a stop might help police develop better training tools and build stronger accountability measures. Independent of any court case, internal monitoring measures can improve hit rates for arrests. In other data-driven contexts, these types of retention and accountability efforts are built into the regulating structure.[325]

(2001) (discussing a proposed law in Illinois that would have required videotaping confessions for certain crimes and made inadmissible confessions not videotaped).

[324] Many police officers are required to document certain police–citizen encounters after the fact. Jeffrey Fagan & Garth Davies, *Street Stops and Broken Windows:* Terry*, Race, and Disorder in New York City*, 28 FORDHAM URB. L.J. 457, 487-88 (2000) (describing the NYPD's use of UF-250 cards to record police–citizen encounters).

[325] *Cf., e.g.*, ABA STANDARDS FOR CRIMINAL JUSTICE: LAW ENFORCEMENT ACCESS TO THIRD PARTY RECORDS § 25-7.1 (3d ed. 2013), *available at* http://www.americanbar.org/content/dam/aba/publications/criminal_justice_standards/third_party_access.authcheckdam.pdf (recommending accountability mechanisms for databases used by law enforcement).

4. Efficiency

A move toward data-driven policing will also improve the efficient use of police resources. In many cases, better information will lead police to focus scarce resources on more serious risks and prevent unnecessary contacts with law-abiding citizens.

The rise of predictive policing signals the beginning of this shift to data-driven tips.[326] Underlying the theory of predictive policing is the idea that areas statistically more likely to have crime should have an additional police presence.[327] The data guide the officer patrol patterns, down to the particular time, date, and location.[328] While not replacing police patrols in other areas, police seek to target the areas identified through data.[329] Police administrators

[326] *See generally* Ferguson, *Predictive Policing*, *supra* note 29, at 265-69 (discussing the use of algorithms to predict crime and allocate law enforcement resources).

[327] *See* Braga et al., *supra* note 225, at 9 ("Criminological evidence on the spatial concentration of crime suggests that a small number of highly active micro places in cities—frequently called 'hot spots'—may be primarily responsible for overall citywide crime trends."); *see also* Joel M. Caplan et al., *Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting*, 28 JUST. Q. 360, 364 (2011) ("While a crime event occurs at a finite place, risk is a continuous dynamic value that increases or decreases intensity and clusters or dissipates in different places over time, even places remote from a crime event."); Shane D. Johnson et al., *Offender as Forager? A Direct Test of the Boost Account of Victimization*, 25 J. QUANTITATIVE CRIMINOLOGY 181, 184 (2009) (positing that the clustering of crimes could be explained by optimal foraging strategies); Shane D. Johnson et al., *Space–Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization*, 23 J. QUANTITATIVE CRIMINOLOGY 201, 203-04 (2007) ("Most criminals commit crimes in areas with which they are already familiar."); Ashley B. Pitcher & Shane D. Johnson, *Exploring Theories of Victimization Using a Mathematical Model of Burglary*, 48 J. RES. CRIME & DELINQ. 83, 85-86 (2011) (discussing two theories that seek to explain the near-repeat phenomenon).

[328] *See generally* Goode, *supra* note 226, at A11 (reporting on anticipatory police deployments in Santa Cruz, California); *Predictive Policing: Don't Even Think About It*, ECONOMIST, July 20, 2013, at 24, 24-26 (describing data-driven police resource allocation); Leslie A. Gordon, *Predictive Policing May Help Bag Burglars—But It May Also Be a Constitutional Problem*, ABA JOURNAL (Sept. 1, 2013, 3:40 AM), http://www.abajournal.com/magazine/article/predictive_policing_may_help_bag_burglars--but_it_may_also_be_a_constitutio/, *archived at* http://perma.cc/J3L3-U9NN (discussing constitutional concerns relating to predictive policing); Kaste, *supra* note 219 (reporting on forward-looking policing strategies used in Seattle and other cities).

[329] *See* Charlie Beck & Colleen McCue, *Predictive Policing: What Can We Learn from Wal-Mart and Amazon About Fighting Crime in a Recession?*, POLICE CHIEF (Nov. 2009), http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=1942&issue_id=112009, *archived at* http://perma.cc/D6DP-UYAS ("Predictive policing allows command staff and police managers to leverage advanced analytics in support of meaningful, information-based tactics, strategy, and policy decisions in the applied public safety environment. As the law enforcement community increasingly is asked to do more with less, predictive policing represents an opportunity to prevent crime and respond more effectively, while optimizing increasingly scarce or limited resources, including personnel.").

have embraced predictive policing because it allows them to allocate resources more efficiently while at the same time reducing crime.

Similarly, collecting data on individuals or groups perceived to be at high risk of entering the criminal justice system allows for a more focused use of police resources. Joint federal and state fusion centers have evolved to tackle gang and gun violence.[330] In these collaborative centers, police, with the help of technology, identify and map individuals by known gang associations, ethnicity, age, race, address, and social connections.[331] In Washington, D.C., one early partnership between federal and local law enforcement resulted in a "gang audit" that "helped identify 136 of the most violent gang/crew members in three of the highest crime areas in D.C."[332] People identified by police as involved in gangs faced targeted interventions, including face-to-face meetings, evictions from public housing, and criminal prosecution.[333] By mapping and targeting only those statistically most likely to be involved in criminal activity, the police attempted to address violence in the community proactively. This was the thinking behind the Chicago "heat list," and the approach has the benefit of focusing resources on those more likely to be involved in crime—whether as perpetrators or victims. While predictive policing practices raise a host of fairness concerns, from an efficiency perspective, recent innovations appear to have been a success.

5. Unexpected Insights

Big data also allows for unexpected insights from the collection of vast amounts of seemingly innocuous information. To package crack cocaine, a drug dealer needs tiny plastic bags and a scale.[334] To fire a gun, a shooter

[330] U.S. DEP'T OF JUSTICE, FUSION CENTER GUIDELINES: DEVELOPING AND SHARING INFORMATION AND INTELLIGENCE IN A NEW ERA F-3 (2006), *available at* http://www.it.ojp.gov/documents/fusion_center_guidelines.pdf (defining a fusion center as "[a] collaborative effort of two or more agencies that provide resources, expertise, and/or information . . . with the goal of maximizing the ability to detect, prevent, apprehend, and respond to criminal and terrorism activity"); *see also* Mimi Hall, *Feds Move to Share Intelligence Faster*, USA TODAY, July 27, 2006, at 3A (reporting that state fusion centers are run by "state police, FBI agents, National Guard, health officials and others").

[331] *Cf., e.g.*, Kelly, *supra* note 152 (describing the rise in use of cell phone information-gathering devices by police departments).

[332] SCOTT DECKER ET AL., PROJECT SAFE NEIGHBORHOODS: STRATEGIC INTERVENTIONS 18 (2007), *available at* https://www.bja.gov/publications/strategic_prob_solving.pdf.

[333] *Id.* at 17-19.

[334] *See* United States v. Dingle, 114 F.3d 307, 309 (D.C. Cir. 1997) (recounting expert testimony on practices used by drug dealers for packaging and distributing crack cocaine).

needs a bullet.[335] To break into a car, a thief needs tools (modern or old fashioned).[336] By tracking the sale of these items, police can recognize patterns and thus identify the criminals making the purchases. Similarly, most major criminal enterprises must launder money and otherwise hide illicit proceeds.[337] Unusual deposits, purchases, or money transfers can allow police to identify money laundering and the people involved.[338]

Incorporating geographic data can reveal patterns of location in an otherwise fluid criminal environment. Knowing where particular crimes occur can allow for more targeted suppression strategies. Big data allows for better tracking of national (or transnational) crimes, including human trafficking, drug smuggling, and credit card fraud.[339] For example, Google and others have partnered with three international antitrafficking nonprofits to track where calls for assistance originate to better map and disrupt human trafficking.[340] Similarly, hospital overdose admissions could reveal

[335] For an interesting story on how data about guns used in violent crime can be tracked and studied, see David S. Fallis, *Tracing Secrets*, WASH. POST, Oct. 24, 2010, at A1, which reports the findings of a *Washington Post* investigation into the sources of guns used in crimes—most notably that one dealer sold more than 2500 guns later recovered by police.

[336] *See, e.g.*, *Today: Rossen Reports* (NBC television broadcast June 5, 2013), *available at* http://www.today.com/news/police-admit-theyre-stumped-mystery-car-thefts-6C10169993 (reporting on a series of car thefts committed using a device that quickly bypasses electronic locks).

[337] Jimmy Gurulé, *The Money Laundering Control Act of 1986: Creating a New Federal Offense or Merely Affording Federal Prosecutors an Alternative Means of Punishing Specified Unlawful Activity?*, 32 AM. CRIM. L. REV. 823, 823 (1995) (describing money laundering as the "lifeblood" of organized crime); *see also Money Laundering Legislation: Hearing on S. 572, S. 1335, and S. 1385 Before the S. Comm. on the Judiciary*, 99th Cong. 30 (1985) (statement of Sen. DeConcini, Member, S. Comm. on the Judiciary) ("Without the means to launder money, thereby making cash generated by a criminal enterprise appear to come from a legitimate source, organized crime could not flourish as it now does.").

[338] *See, e.g.*, Richard K. Gordon, *Losing the War Against Dirty Money: Rethinking Global Standards on Preventing Money Laundering and Terrorism Financing*, 21 DUKE J. COMP. & INT'L L. 503, 527-28 (2011) (describing the "red flags" used by the Treasury Department's financial intelligence unit, FinCEN, to identify money laundering).

[339] *See, e.g.*, Philip K. Chan et al., *Distributed Data Mining in Credit Card Fraud Detection*, IEEE INTELLIGENT SYSTEMS, Nov.–Dec. 1999, at 67, 68 (providing technical details of specific credit card fraud identification algorithms); Scott R. Peppet, *Prostitution 3.0?*, 98 IOWA L. REV. 1989, 2039-40 (2013) (suggesting data with which to estimate the likelihood that a prostitute is a victim of human trafficking); Peter P. Swire, *Privacy and Information Sharing in the War on Terrorism*, 51 VILL. L. REV. 951, 964 (2006) (discussing the "out of pattern" system for identifying credit card fraud).

[340] Bernhard Warner, *Google Turns to Big Data to Unmask Human Traffickers*, BLOOMBERG BUSINESSWEEK (Apr. 10, 2013), http://www.businessweek.com/articles/2013-04-10/google-turns-to-big-data-to-unmask-human-traffickers, *archived at* http://perma.cc/3CSC-RDUJ ("The [Google-led] alliance . . . means the three anti-trafficking networks . . . will share data on where the emergency phone calls are originating, the ages of the victims, their home countries, and the types of criminal activities they have been forced into. . . . [T]he agencies will be able to crunch data

drug use patterns. Social media trends may reveal clues about gang activities,[341] prostitution services,[342] or cybercrime.[343]

Patterns of crime can also reveal the locations of criminal actors. Police can link certain getaway routes to robbery hotspots.[344] Locations of gunshots can reveal shifting gang-related turf borders.[345] Social services visits to monitor "stay-away orders" can reveal potential locations of future domestic violence.[346] Even the type of alcohol sold at stores and restaurants can correlate with the rate of violent crime in a neighborhood.[347] These insights can help police investigate and prevent crime and would not have been easily observed before the advent of big data.

like this in real time to detect crime trends that they can then share with police and policymakers to help protect victims.").

[341] *See, e.g.*, Ben Austen, *Public Enemies: Social Media Is Fueling Gang Wars in Chicago*, WIRED (Sept. 17, 2013, 6:30 AM), http://www.wired.com/2013/09/gangs-of-social-media/, *archived at* http://perma.cc/3L5H-L2M2 (describing escalating gang tensions via Twitter and YouTube).

[342] *See, e.g.,* Erica Fink & Laurie Segall, *Pimps Hit Social Networks to Recruit Underage Sex Workers*, CNNMONEY, (Feb. 27, 2013, 7:30 AM), http://money.cnn.com/2013/02/27/technology/social/pimps-social-networks, *archived at* http://perma.cc/S4BU-LEUK (reporting on the use of Facebook and other social media sites to lure victims into becoming sex workers).

[343] *See generally Online Privacy, Social Networking, and Crime Victimization: Hearing Before the Subcomm. on Crime, Terrorism, & Homeland Sec. of the H. Comm. on the Judiciary*, 111th Cong. 5-12 (2010) (statement of Gordon M. Snow, Asst. Dir., FBI) (discussing ways in which cybercriminals use social media to deceive victims).

[344] JENNIFER BACHNER, IBM CTR. FOR THE BUS. OF GOV'T, PREDICTIVE POLICING: PREVENTING CRIME WITH DATA AND ANALYTICS 15-16 (2013), *available at* http://www.businessofgovernment.org/sites/default/files/Predictive%20Policing.pdf. (suggesting that criminals prefer "areas with desirable escape routes," including "[a]reas in close proximity to features such as interstate highways, bridges, and tunnels").

[345] *See* Andras Petho, David S. Fallis & Dan Keating, *Acoustic Sensors Reveal Hidden Depth of Gun Use in D.C.*, WASH. POST, Nov. 2, 2013, at A1 (describing data from the District of Columbia's acoustic "ShotSpotter" system, which had identified 39,000 separate instances of gunfire, many of which were clustered geographically).

[346] Goldstein, *supra* note 294, at A1 (discussing efforts by the NYPD to reduce domestic violence).

[347] Robert Lipton et al., *The Geography of Violence, Alcohol Outlets, and Drug Arrests in Boston*, 108 AM. J. PUB. HEALTH 657, 661 (2013) (suggesting "a positive relationship between violent crime and the presence of package stores," but "a negative relationship between violent crime and the presence of restaurants selling beer and wine."); *see also* Press Release, Univ. of Mich. Health Sys., Could a Computer on the Police Beat Prevent Violence? (Feb. 18, 2013), *available at* http://www.uofmhealth.org/news/archive/201302/could-computer-police-beat-prevent-violence ("Results from the study indicate that types and densities of alcohol outlets were directly related to violent crimes despite the fact that alcohol outlets are typically viewed as locations in which other population or environmental factors, such as poverty or prostitution, relate to the violence.").

B. *The Negatives of Big Data Suspicion*

While big data offers much promise, big data–driven policing also has potential negative consequences. This Section outlines a few representative concerns.

1. Bad Data

A system based on data requires accurate, up-to-date information.[348] One concern with a vast, ever-growing, networked data system is that the quality controls on shared data are almost nonexistent.[349] Police may rely on existing data without any knowledge of how the data was collected or whether mechanisms exist to ensure its accuracy. Data problems have emerged even within locally controlled systems[350] and certainly arise when jurisdictions share information.[351] Reputed "gang lists" used by police have been shown to be inaccurate.[352] Arrest reports can be inaccurate or erroneous

[348] Cope, *supra* note 24, at 193 ("Data quality affected the development of analysis. Analysts frequently found crucial details missing from intelligence reports for their products.").

[349] *See, e.g.*, Eric J. Mitnick, *Procedural Due Process and Reputational Harm: Liberty as Self-Invention*, 43 U.C. DAVIS L. REV. 79, 126 (2009) (noting that while most databases are supposed to be subject to quality control, "[i]n reality . . . , the evidence is overwhelming that the control measures currently in place regularly fail, either due to lack of resources, skill, or because they are simply neglected"); Wright, *supra* note 121, at 122 (finding quality control lacking in one database where no reports were questioned by superiors; the officers making some of the reports had no gang experience, and there were no reviews for accuracy).

[350] *See* Jeff Morganteen, *What the CompStat Audit Reveals About the NYPD*, N.Y. WORLD (July 3, 2013), http://www.thenewyorkworld.com/2013/07/03/compstat/, *archived at* http://perma.cc/K4ZP-KR4L ("The outside audit . . . not only confirmed that such data manipulation takes place but found several weak points in the ways the department tracks and uncovers it."); *see also* DAVID N. KELLEY & SHARON L. MCCARTHY, THE REPORT OF THE CRIME REPORTING REVIEW COMMITTEE TO COMMISSIONER RAYMOND W. KELLY CONCERNING COMPSTAT AUDITING 47 (2013), *available at* http://www.nyc.gov/html/nypd/downloads/pdf/public_information/crime_reporting_review_committee_final_report_2013.pdf ("[T]he patterns of the misclassified reports support in some measure the anecdotal accounts . . . that certain types of incidents may be downgraded as a matter of practice in *some* precincts.").

[351] Herring v. United States, 555 U.S. 135, 155 (2009) (Ginsburg, J., dissenting) ("The risk of error stemming from these databases is not slim. Herring's *amici* warn that law enforcement databases are insufficiently monitored and often out of date. Government reports describe, for example, flaws in NCIC databases, terrorist watchlist databases, and databases associated with the Federal Government's employment eligibility verification system." (footnotes and citation omitted)).

[352] *See, e.g.*, Mitnick, *supra* note 349, at 126; Wright, *supra* note 121, at 129 ("In sum, gang databases appear to be riddled with factual inaccuracies, administrative errors, lack of compliance with departmental guidelines, and lack of oversight.").

but remain in public and private databases.[353] The FBI's own files—used for millions of background checks—reportedly contain hundreds of thousands of errors.[354] Worse, there is no simple mechanism to clear the bad data from a web of networked systems all sharing the same errors.[355]

Adding private, third-party sources of information only compounds the problem. CBS News's 60 Minutes reported that "as many as forty million Americans have a mistake on their credit report. Twenty million have significant mistakes."[356] These are the same credit report datasets that underlie many commercial big data systems. Both discovering and correcting mistakes is difficult; it requires knowledge of the error and the wherewithal to change it. Police agents accessing records, however, would have no knowledge that an error existed—or even necessarily a way to check the accuracy of the data. Mistakes can occur at any point in the process from collection to entry to analysis. In addition, data can grow stale. Typographical errors can lead to erroneous linkages.[357] These mistakes can have real consequences on individual liberty. As Justice Ginsburg warned:

> Inaccuracies in expansive, interconnected collections of electronic information raise grave concerns for individual liberty. The offense to the dignity of the citizen who is arrested, handcuffed, and searched on a public street simply because some bureaucrat has failed to maintain an accurate computer data base is evocative of the use of general warrants that so outraged the authors of our Bill of Rights.[358]

[353] Roberto Concepción, Jr., *Need Not Apply: The Racial Disparate Impact of Pre-Employment Criminal Background Checks*, 19 GEO. J. ON POVERTY L. & POL'Y 231, 246-48 (2012) (highlighting the high cost of false positives in pre-employment queries of criminal records databases).

[354] *See* Ylan Q. Mui, *Use of FBI Database in Hiring Raises Concerns*, WASH. POST, July 30, 2013, at A1 (discussing a report by the National Employment Law Project on errors in FBI background checks).

[355] *See, e.g.*, Anita Ramasastry, *Lost in Translation? Data Mining, National Security and the "Adverse Inference" Problem*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 757, 775-76 (2006) (discussing reports of errors and inaccuracies in credit reports); Tal Z. Zarsky, *Governmental Data Mining and Its Alternatives*, 116 PENN ST. L. REV. 285, 298 (2011) (discussing the problem of errors in data mining processes).

[356] *60 Minutes: 40 Million Mistakes: Is Your Credit Report Accurate?* (CBS television broadcast Feb. 10, 2013), *available at* http://www.cbsnews.com/8301-18560_162-57567957/credit/.

[357] *Cf.* Wayne J. Pitts, *From the Benches and Trenches: Dealing with Outstanding Warrants for Deceased Individuals: A Research Brief*, 30 JUST. SYS. J. 219, 220 (2009) (describing a study that discovered numerous errors in a warrant database, including incorrect social security numbers, inaccurate names, and "illogical birth dates," and noting that "none of the[] issues are surprising or unusual given the nature of the population being tracked").

[358] Herring v. United States, 555 U.S. 135, 155-56 (2009) (Ginsburg, J., dissenting) (internal quotation marks omitted).

The lack of transparency in these data systems only increases the chance of error. Police systems are usually restricted to authorized police users. Private companies, seeking commercial gain, have little incentive to reveal the workings of proprietary systems or the data thereby collected. No agency has the responsibility to audit the growing governmental and commercial big data network. While the Federal Trade Commission has promised to monitor private big data companies,[359] it has little ability to examine the data itself and has no role in oversight of law enforcement use of the data. Though oversight institutions do exist (including courts, congressional committees, and independent agencies),[360] the volume of information at issue prevents these groups from examining the quality of the data or the magnitude of the errors.[361] Without transparency, there can be little hope for accountability to ensure that data systems will be sufficiently reliable to justify altering constitutional rights.[362] In short, big data suspicion may be based on bad data.

[359] *See* Brendan Sasso, *FTC Chief Targets Firms with Vast Databases*, HILL (Aug. 19, 2013, 9:12 PM), http://thehill.com/policy/technology/317729-ftc-chief-targets-firms-with-vast-databases, *archived at* http://perma.cc/8HTB-SE4W (reporting that the head of the FTC stated that the agency "will use its power to punish deceptive business practices [and] to crack down on firms that fail to live up to their own promises about how they will use their customers' data"). *See generally* FTC, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY 46-56 (2014), *available at* http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf (presenting findings of an FTC study of large data brokers and recommending reforms).

[360] For example, the House Select Committee on Intelligence and the Senate Select Committee on Intelligence have legislative oversight of the intelligence agencies. The House Committee on the Judiciary, the Senate Committee on the Judiciary, the House Committee on Homeland Security, the Senate Committee on Homeland Security and Governmental Affairs, and others have oversight of domestic surveillance. Independent agencies such as the Privacy and Civil Liberties Oversight Board have general oversight. The Foreign Intelligence Surveillance Court provides some judicial oversight. General counsels and inspectors general add additional layers of protection.

[361] For example, the Foreign Intelligence Surveillance Court released a redacted opinion offering insight into problems with overcollection of phone records by the National Security Agency. *See* [Redacted], [Redacted], 2011 WL 10945618 (FISA Ct. Oct. 3, 2011). In its October 2011 opinion, the court revealed that it could review only samples of the NSA-collected data due to the incredible number of search queries and volume of data involved with the NSA's operations. *See id.* at *10; *see also In re* Order Requiring Production of Tangible Things From [Redacted], 2013 WL 5741573, at *10-14 (FISA Ct. Aug. 29, 2013) (No. BR 13-109) (setting guidelines for review of NSA metadata-related surveillance programs); MAJORITY STAFF, SENATE COMM. ON COMMERCE, SCI. & TRANSP., 113TH CONGRESS., A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE, AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES (2013), *available at* http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577 (providing an example of a congressional investigation into data brokers and the collection of personal information).

[362] *See* Solove, *Data Mining*, *supra* note 177, at 359 ("Another key issue regarding the liberty side of the balance is transparency—the degree of openness by which a particular security measure

2. False Positives

Even assuming "good data," big data reasonable suspicion will result in false positives whereby police stop innocent people.[363] This burden will fall most heavily on individuals who have some criminal history, but who are not currently engaging in criminal activity.[364] Predictive analytics will suggest suspicion based on an identified correlation, but such suspicion will often be unfounded. Perhaps a license plate reader will place the car of a convicted burglar within a predicted burglary hotspot, which also happens to be next to the convicted burglar's grandmother's house. Police might stop the suspected burglar solely because of this correlation. One can imagine that those individuals who find themselves on a "bad guy list" will be marked for more than their fair share of borderline suspicious stops.[365]

Big data suspicion creates the real concern that certain individuals by virtue of their past criminal activities will always be at risk to be stopped. Those with lengthy criminal records or gang associations may be stopped because of who they are and not what they are doing. Prior police contacts will become the digital "scarlet letter" marking certain people in a community as suspicious.[366]

Over the past several decades, poor people and people of color have had disproportionate contact with the criminal justice system.[367] If these contacts become data points that can be used in a reasonable suspicion analysis, then these data may become proxies for race or class (with similar

is carried out. Transparency is essential to promote accountability and to provide the public with a way to ensure that government officials are not engaging in abuse.").

[363] *Cf.* Taslitz, *supra* note 311, at 10 ("Any concept of reasonable suspicion . . . that tolerates massive false negative rates—frequent invasions of privacy, property, and locomotive rights that ensnare the apparently innocent—is a flawed conception. The costs imposed on communities and individuals become great, while little in the way of crime-control efforts is achieved.").

[364] *See supra* Part III.

[365] Of course, these individuals might also be targeted without a big data–inspired list.

[366] *See* David Wolitz, *The Stigma of Conviction: Coram Nobis, Civil Disabilities, and the Right to Clear One's Name*, 2009 BYU L. REV. 1277, 1316 (arguing that a criminal conviction is a "uniquely stigmatizing piece of information" and that it disproportionately affects a person's reputational profile).

[367] *See* Robin Walker Sterling, *Raising Race*, CHAMPION, Apr. 2011, at 24, 24-25 ("The criminal justice system has exploded outside of the prison walls, as well. As of 2009, the number of people under criminal justice supervision—including those who are in jail, in prison, on probation, and on parole—totaled 7.2 million people. In a dismaying parallel to incarceration rates, people of color are also overrepresented among arrestees, probationers, and parolees. There are more African Americans under correctional control today than were enslaved in 1850. . . . With numbers like these, it is clear that this overrepresentation of minorities in the criminal justice system, or disproportionate minority contact (DMC), is one of the major human rights violations of our time." (footnotes omitted)).

effect). For example, the ACLU's recent national study on marijuana arrests demonstrates that African Americans are more likely to be arrested for marijuana than whites, despite equivalent usage rates.[368] Thus, more data has been collected about minority marijuana possession, even though whites commit the crime at the same rate. If data are collected only about certain classes of people, then those people are more likely to become future targets of suspicion simply because of the initial selection bias. Thus, important questions remain about who collects, interprets, and chooses the big data to study.[369]

Worse, like other quantitative systems used for decisionmaking, big data–based predictive policing will appear to be objective and fair when it may in fact reflect subjective factors and structural inequalities. Just as we have credit ratings that allow lenders to predict future creditworthiness, police could develop "criminal ratings" to predict future criminal proclivity.[370]

Similarly, data can lead us to believe our own worst instincts.[371] If published data demonstrate a higher arrest rate for people of color, then this information may well influence discretionary decisions about who to stop.[372] Implicit bias and confirmation bias will result in police seeing what they have been told to see, even if it is not actually occurring.[373] Implicit bias involves unconscious prejudices that influence individuals making discretionary

[368] ACLU, THE WAR ON MARIJUANA IN BLACK AND WHITE 17-22 (2013), *available at* https://www.aclu.org/files/assets/aclu-thewaronmarijuana-rel2.pdf (reporting that blacks are roughly four times more likely to be arrested for marijuana possession than whites despite similar usage rates); *see also* Steven Nelson, *ACLU Marijuana Study: Blacks More Likely to be Busted*, U.S. NEWS & WORLD REP. (June 4, 2013, 5:26 PM), http://www.usnews.com/news/newsgram/articles/2013/06/04/aclu-marijuana-study-blacks-more-likely-to-be-busted, *archived at* http://perma.cc/W5DS-ZBUG (reporting on the ACLU study).

[369] *See* Cohen, *supra* note 7, at 1922 ("It is beyond serious question that the techniques that comprise Big Data offer vitally important strategies for promoting human flourishing in an increasingly complex, crowded, and interdependent world. But those techniques cannot themselves decide which questions to investigate, cannot instruct us how to place data flows and patterns in larger conceptual or normative perspective, and cannot tell us whether and when it might be fair and just to limit data processing in the service of other values.").

[370] Thank you to the discussants at Northeastern University School of Law's Legal Scholarship 4.0 conference for developing the concept of "criminal ratings."

[371] Thank you to the discussants at the criminal law professor workshop at the Washington College of Law, American University, for developing this argument.

[372] *Cf.* Taslitz, *supra* note 311, at 44-45 (discussing the potential for extrapolation from past experience despite insufficient information).

[373] *See* Tracey G. Gove, *Implicit Bias and Law Enforcement*, POLICE CHIEF, Oct. 2011, at 44, 50 ("The study of implicit bias has important implications for police leaders. Police officers are human and, as the theory contends, may be affected by implicit biases just as any other individual. In other words, well-intentioned officers who err may do so not as a result of intentional discrimination, but because they have what has been proffered as widespread human biases.").

decisions.[374] This can result in unequal outcomes for similarly situated individuals.[375] Implicit bias inevitably exists in the ordinary course of police activities, but is even more damaging when combined with confirmation bias: "the tendency to bolster a hypothesis by seeking consistent evidence while minimizing inconsistent evidence."[376] Thus, an officer conditioned to believe that a particular type of person may be more likely to commit a criminal act will likely see that person through the lens of suspicion. By providing the information to confirm this suspicion, big data will make it easier for police to justify a stop. Even more dangerously, an officer with discriminatory animus may be able to justify a knowingly unconstitutional stop using an aggregation of otherwise innocent data.[377]

This risk demonstrates how suspicions about past criminal actions can all too easily morph into suspicions about current criminal activity. It highlights the importance of requiring a nexus between the suspected criminal and the suspected criminal activity. It also highlights the dangers of how big data can target certain populations based on correlations with possible criminal activity, rather than causation from real criminal activity. Justification to stop these individuals—marked by big data—will be too easily met, undermining the individualized and particularized protections in the Fourth Amendment.

3. Shifting Power Balance

The Constitution establishes a power-sharing relationship between citizens and the government. The Fourth Amendment, like other parts of the Bill of Rights, represents a check on government power.[378] The probable cause standard, and to a lesser extent, the reasonable suspicion standard, limits the actions of government agents. Big data, by weakening the reasonable

[374] Richardson, *supra* note 312, at 271-72.

[375] *See* Mary Fan, *Street Diversion and Decarceration*, 50 AM. CRIM. L. REV. 165, 192 (2013) ("A rich body of literature has documented how implicit biases—negative perceptions of minorities that may unconsciously lurk despite best intentions—impact the judgment of an array of actors, such as police, prosecutors, and jurors.").

[376] Barbara O'Brien, *Prime Suspect: An Examination of Factors That Aggravate and Counteract Confirmation Bias in Criminal Investigations*, 15 PSYCHOL. PUB. POL'Y & L. 315, 315 (2009); *see also id.* at 318 (noting that "[p]olice investigators are also prone to confirmation bias").

[377] *See* Richard Winton et al., *LAPD to Build Data on Muslim Areas*, L.A. TIMES, Nov. 9, 2007, at A1 (describing a police initiative to identify areas "at-risk" for terrorist activities based on ethnicity); Richard Winton et al., *Outcry Over Muslim Mapping*, L.A. TIMES, Nov. 10, 2007, at A1 (same).

[378] *See* United States v. Martinez-Fuerte, 428 U.S. 543, 554 (1976) (noting that the purpose of the Fourth Amendment is to protect against "arbitrary and oppressive interference by enforcement officials with the privacy and personal security of individuals").

suspicion standard, restructures this relationship, with police gaining more power and citizens losing a measure of liberty. Though citizens are complicit in giving up much of that information to third parties, the government now has more information and can use that information to investigate.[379] Possessing the information, and letting citizens know the government possesses the information, might alone shape individual choices. Citizens may be more reluctant to associate with certain groups, participate in certain activities, or even take political stances because of the information the government knows about their private lives.[380] The collection of the information may be as threatening as its potential use.[381] Privacy scholars have ably addressed this issue as a consequence of the new big data world.[382] And, for police–citizen encounters on the street, this informational power could alter behavior.

V. TENTATIVE SOLUTIONS

This Article is an attempt to sketch the potential impact of big data information on a small data reasonable suspicion doctrine. It has exposed real distortions and potential vulnerabilities, as well as clear advantages, that arise from this technological innovation. The debate over big data is noisy and contested, perhaps revealing its immature state of development. This last Part addresses tentative solutions in this changed landscape.

[379] *See* Jed Rubenfeld, *The End of Privacy*, 61 STAN. L. REV. 101, 118 (2008) ("[T]he government's law enforcement power is unique. . . . The ability of government to intrude, monitor, punish, and regulate is greater than that of private actors by many orders of magnitude. But more than this, the state has a right and duty to intrude into people's lives that private parties do not. . . . But precisely because the state's law enforcement power gives it a license to intrude into our homes and lives in ways that private parties cannot, the state poses dangers to a free citizenry that private parties do not.").

[380] *See* Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1946-48 (2013) (discussing the dangers to intellectual privacy from surveillance).

[381] Harold J. Krent, *Of Diaries and Data Banks: Use Restrictions Under the Fourth Amendment*, 74 TEX. L. REV. 49, 60 (1995) ("Particularly in light of new technology, privacy is threatened as much by what law enforcement authorities do with information as by the original acquisition itself.").

[382] *See generally* Jane Yakowitz Bambauer, *The New Intrusion*, 88 NOTRE DAME L. REV. 205, 257-73 (2012) (suggesting that expanded tort law could help address violations of data privacy, including harm from subsequent disclosures); Andrew Jay McClurg, *Bringing Privacy Law Out of the Closet: A Tort Theory of Liability for Intrusions in Public Places*, 73 N.C. L. REV. 989, 1076 (1995) (discussing the demise of the tort of public disclosure of private facts and the resulting gap in privacy law); Paul Ohm, *The Fourth Amendment in a World Without Privacy*, 81 MISS. L.J. 1309, 1334-35 (2012) (arguing that in a world of vanishing privacy, a view of the Fourth Amendment as addressing a balance of power between government and the people is more appropriate than limiting it strictly to a right to privacy).

A. *Doctrinal Changes*

Initially, we could simply require a more stringent standard of reasonable suspicion. To be clear, the Supreme Court has been steadfast in articulating that it has no intention of quantifying—or even clarifying—the standard, instead recognizing that police officers operate within "the factual and practical considerations of everyday life on which reasonable and prudent men, not legal technicians, act."[383] But if the underlying logic of a standard based on unknown suspects has been upended by all-knowing technology, then perhaps the standard needs to be altered. The "factual and practical considerations" are no longer limited to small data observations, so why shouldn't the factual and practical considerations include requiring more data now easily available? One could imagine that in the context of big data predictive suspicion, where police can more easily access information, courts might require more information to satisfy the reasonable suspicion standard.[384]

In practical terms, this could result in a different standard of reasonable suspicion when big data information is used. If big data makes more information available with relatively little effort, then big data should be required to be part of the reasonable suspicion calculus. This would include both potentially suspicious and potentially exculpatory information. After all, if the original justification for reasonable suspicion arose out of a situation necessarily limited by practicality (unknown suspect, potentially imminent crime, limited information), then better information about a known suspect suggests a different standard should be used.[385]

B. *Big Data Changes*

If changing the standard of reasonable suspicion is not a realistic option, then perhaps the solution lies in the nature of big data itself. At its core, big data encourages a heightened focus on (1) statistical analysis, (2) geospatial

[383] Ornelas v. United States, 517 U.S. 690, 695 (1996) (quoting Illinois v. Gates, 462 U.S. 213, 231 (1983)).

[384] While beyond the scope of this Article, there may be an unarticulated taxonomy of reasonable suspicion that applies differently in different circumstances based on the type of crime at issue. It may be that an analysis of big data's impact simply reveals this unacknowledged truth.

[385] One reason the *Terry* court seemed willing to create a standard less than probable cause was because practicality demanded it. Detective McFadden simply could not obtain any other information about John Terry without stopping him. If McFadden let the suspects go, he might never have been able to identify them. Big data suspicion changes one part of that calculus—information is now available to police without leaving the scene. Information about the suspect may alter the level of suspicion, even in potentially violent and imminent crimes. In addition, identifying suspects for certain possession offenses may allow police to monitor the suspects without necessarily stopping them at that moment.

and temporal analysis, and (3) link analysis. Adapting these three methods of analysis to big data suspicion on the streets might provide some precision and, thus, some limitation for an otherwise malleable standard.

1. Statistical Analysis

First, big data not only involves making predictions, but also quantifying the likelihood that these predictions will come true. The statistical analysis behind the predictive analytics provides actual, observable numbers. For example, a predictive policing algorithm might report a 2.06% chance of a burglary on a particular block.[386] A pattern match might provide police with a percentage likelihood that a particular suspect associates with another individual. In these cases, if part of the predictive suspicion of an individual comes from big data, courts could require (when available) a numerical prediction of accuracy.[387] This quantification process would necessarily result in courts having to confront the ultimate question of how "certain" a police officer must be to constitute reasonable suspicion.[388] This has been a forbidden discussion in the courts, and the possibility of such a discussion would likely forestall any adoption of this proposal. Yet big data creates the promise of using quantitative and empirical evidence to refine what has ordinarily been a quasi-subjective judgment. Courts and police could adapt this hallmark of the big data revolution to sharpen the reasonable suspicion analysis. Thus, a court might find a likelihood of 2.06% to be insufficient in the reasonable suspicion analysis but a likelihood of 20.06% to be enough. Even if courts were unwilling to establish a particular percentage threshold as reasonable suspicion, the quantified likelihood could still be included in the totality of circumstances calculation.

Related to this threshold determination would be the more difficult question of whether the threshold should change depending on the type of crime at issue. Should a 10% predictive judgment be enough for a murder investigation but not a drug crime? Should the scale be a sliding one based

[386] Kalee Thompson, *The Santa Cruz Experiment*, POPULAR SCI., Nov. 2011, at 38, 40 (describing a prediction in Santa Cruz, California, that on "Linden Street, where, the statistics reveal, there is a 2.06 percent chance of a crime happening today, and 3:1 odds that a crime, should it occur, will be a home break-in versus an auto theft").

[387] *See* Ferguson & Bernache, *supra* note 68, at 1607-22 (discussing cases in which courts have handled evidence of and sought to define "high-crime" areas).

[388] *See, e.g.*, Max Minzner, *Putting Probability Back into Probable Cause*, 87 TEX. L. REV. 913, 958 (2009) (arguing that judges should be allowed to consider success rates when dealing with probable cause claims and warrants); Taslitz, *supra* note 315, at 202-04 (discussing the qualitative and quantitative requirements the Supreme Court uses for anticipatory warrants).

on risk of harm to society?[389] Again, going back to first principles, *Terry* involved a potential armed robbery. This may be different than a potential shoplifting offense. Risk assessment mechanisms already exist that use an adjusted actuarial model that starts with a fixed percentage and then adjusts based on other factors.[390] While not perfectly aligned, this type of model may prove useful in designing a quantifiable percentage that courts would be comfortable adopting.

At this early stage of the development of big data, any normative argument about how statistical precision should be incorporated into a police officer's calculations will ultimately be unsatisfying. The algorithms do not yet exist that would project numerical likelihoods. While one could demand a specific numerical target, we simply do not know how big data technology will be integrated in everyday policing, nor do we know how data will influence police officers' discretionary decisions. The focus of this Article is to highlight the existence of big data policing and, instead of ignoring it or pretending it will not affect police officers on the street, make courts and scholars aware of the constitutional concerns.

2. Precision of Place and Time

Second, in evaluating big data's influence, courts could focus on the precision of the data, in terms of place and time, to observe how the data correlate with the predicted suspicion. Big data's value as a predictive tool involves its ability to drill down to specific factors that identify a specific person, at a specific time, at a specific place. Because all predictions might come true at a certain level of abstraction, big data also poses risks. An officer who sees a suspect on the street outside the jewelry store and looks up the suspect's criminal history, which includes several theft convictions, would at one level have persuasive information to consider in the totality of circumstances. If after further examination, however, the officer found that the thefts were exclusively downloading digital music, the data would not support suspicion of a jewelry store robbery. Or, if the timing of the prior thefts were several years old or in another state, the correlation might not be strong. Reliance on big data—because it can provide granular information on particular events or linkages—should also necessitate a requirement of

[389] *See, e.g.*, Christopher Slobogin, *A Jurisprudence of Dangerousness*, 98 NW. U. L. REV. 1, 50-58 (2003) (discussing proportionality and consistency principles).

[390] *See, e.g.*, Christopher Slobogin, *Dangerousness and Expertise Redux*, 56 EMORY L.J. 275, 277, 288-93 (2006) (highlighting the advantages and disadvantages of using actuarial prediction techniques and clinical techniques).

precision. A requirement of geospatial and temporal precision would provide some measure of protection.

As a related point, if the reasonable suspicion standard incorporated an "imminence requirement" such that the predicted crime must be imminent or occurring, the general suspicion created by big data would be limited. Such a requirement would conflict with *Hensley* (which allows for reasonable suspicion of already completed crimes),[391] but would provide a counterweight to the power of general suspicion of some unstated "criminal activity." Police officers would need to be able to articulate the particular crime that was imminent based on the data or be foreclosed from using the information. Grafting on an imminence requirement could offer more protection within the existing reasonable suspicion analysis.

Requiring a heightened level of technological precision is consistent with general Fourth Amendment principles requiring particularity and individualization[392] and preventing arbitrary invasions of individual security.[393] In addition, the availability of big data sources now allows for this level of specificity. The question for reviewing courts will be whether to require it. If it were possible to obtain more particularized information from available sources, and officers choose not to obtain that information instead relying on generalized data, then judges, like they routinely do with warrants, could deny the use of generalized data in their reasonable suspicion findings. Judges will be in the best position to make these decisions, and the more precision that judges demand, the more incentive there will be for police to generate precise information to support their suspicion.

3. Limited Link Analysis

Finally, predictive suspicion must embrace the cautions arising from link analysis, which examines the connections between data points.[394] In the context of criminal activity, link analysis refers to the connections between criminals and between crimes. Though link analysis can help find matches in patterns, identify anomalies where known patterns are absent, and discover new patterns, it remains fundamentally overinclusive. There will

[391] United States v. Hensley, 469 U.S. 221, 227-29 (1985).

[392] For a wonderfully insightful analysis of how individualization has been analyzed in the Fourth Amendment context, see Bambauer, *supra* note 303.

[393] *See* Andrew Guthrie Ferguson, *Personal Curtilage: Fourth Amendment Security in Public*, 55 WM. & MARY L. REV. 1283, 1327-31 (2014) (discussing the importance of protecting against arbitrary government action as part of the "security" focus of the Fourth Amendment).

[394] Richard Gordon, *Terrorism Financing Indicators for Financial Institutions in the United States*, 44 CASE W. RES. J. INT'L L. 765, 779 (2012) ("Link analysis explores associations within collections of data.").

always be links that demonstrate nothing suspicious. Correlations do not prove causation.

In order to overcome uncertainty in data, analysts have to go beyond direct links (sometimes several links or "hops" out) and instead look for approximations in identifying data. This indirectness invites error in identification. People have the same names.[395] People are related to criminals without being criminals themselves. People forsake the criminal life.[396] In a wired world, people are more closely connected than we think. Researchers have validated the "six degrees of separation" theory in the information age.[397] One study of Facebook users showed that the "average number of acquaintances separating any two people in the world was not six but 4.74."[398] Thus, a link analysis that goes out three "hops" can cast a very wide net, accidentally capturing many people who are only suspicious by this loose, associative relationship.

In response, predictive suspicion based on link analysis must demand a tighter connection between suspects. Courts must not blindly accept associational suspicion created by an algorithm stretched to link two people. Instead, the link must be tightly controlled. Currently, no rules govern these links, but in order to prevent the use of associational correlations to determine suspicion, some tighter limits must be required.

[395] *See, e.g.*, U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-06-1031, TERRORIST WATCH LIST SCREENING: EFFORTS TO HELP REDUCE ADVERSE EFFECTS ON THE PUBLIC 2 (2006), *available at* http://www.gao.gov/new.items/d061031.pdf ("Because terrorist watch list screening involves comparisons based on personal-identifying information such as names and dates of birth, there is potential to generate misidentifications—given that two or more persons, for example, may have the same or similar names."); McIntire, *supra* note 120, at A1 (describing how some targets of the government's terrorist watch list are "victims of errors in judgment or simple mistaken identity").

[396] *See generally, e.g.*, R. DWAYNE BETTS, A QUESTION OF FREEDOM: A MEMOIR OF SURVIVAL, LEARNING, AND COMING OF AGE IN PRISON (2009) (recounting the author's eight years in Virginia prisons and the effect it had on his life); SHON HOPWOOD WITH DENNIS BURKE, LAW MAN: MY STORY OF ROBBING BANKS, WINNING SUPREME COURT CASES, AND FINDING REDEMPTION (2012) (detailing the life story of a bank robber-turned jailhouse lawyer); Lonnae O'Neal Parker, *From Inmate to Mentor, Through the Power of Books*, WASH. POST, Oct. 2, 2006, at A1 (recounting the story of Dwayne Betts, who served time before reforming and starting a book club for youth).

[397] *See* David Smith, *Proof! Just Six Degrees of Separation Between Us*, OBSERVER, Aug. 3, 2008, at 7 (reporting that Microsoft studied email communications and found an average of 6.6 degrees of separation between any two people).

[398] John Markoff & Somini Sengupta, *Separating You and Me? 4.74 Degrees*, N.Y. TIMES, Nov. 21, 2011, at B1.

CONCLUSION

The question that opened this Article was "whether a Fourth Amendment stop can be predicated on the aggregation of specific and individualized, but otherwise noncriminal, factors."[399] In a big data world, the answer appears to be—perhaps troublingly—yes, if those particularized factors can be connected to observed actions. For those who are concerned that the reasonable suspicion standard has already allowed for overly aggressive policing, discriminatory policing, and unaccountable policing, this conclusion will only raise the level of concern. At the same time, more accurate data may well prevent many of the "rough justice" tactics that are based on class, race, or age profiling and that have nothing to do with the actual individual involved.

In either case, the rise of big data is only just beginning. The search for new data sources and connections has just commenced, and as society's technological capabilities improve, the law must similarly evolve. Police officers on patrol in 2015 may not be able to immediately scan a crowd to reveal identities, but that technology is coming.[400] As with many technological innovations, the law has lagged behind. The concerns raised in this Article will soon be addressed by courts forced to confront how to evaluate reasonable suspicion in a big data world. Perhaps this change will involve new interpretations of the reasonable suspicion standard, or perhaps courts (or legislatures) will adopt wholly new legal standards. But, the law will have to adapt because the current small data reasonable suspicion standard cannot survive the big data era.

[399] *See supra* p. 330.

[400] Rushin, *supra* note 154, at 288 ("Facial recognition software has already been combined with video surveillance and used by law enforcement to identify potential suspects amongst large crowds.").