NORTH CAROLINA JOURNAL OF LAW & TECHNOLOGY
VOLUME 16, ISSUE 2: JANUARY 2015

WHEN BIG DATA MEETS BIG BROTHER: WHY COURTS SHOULD APPLY *UNITED STATES V. JONES* TO PROTECT PEOPLE'S DATA

Brad Turner*

In an age when people's lives are constantly tracked, recorded, analyzed, and shared by private parties, the Third-Party Doctrine, which holds that "information knowingly exposed to private parties is unprotected by the Fourth Amendment," now threatens to swallow whole the privacy guaranteed by the Fourth Amendment. This Article suggests courts adopt the Klayman v. Obama *approach and hold that the Fourth Amendment's protections apply to government acquisitions of Big Data. More specifically, courts should follow Justice Samuel Alito's reasoning in* United States v. Jones *to hold that government acquisitions of Big Data are searches subject to the reasonableness requirements of the Fourth Amendment. Surely, if the government's collection of a person's GPS data in* Jones *was intrusive enough to constitute a search, then so too should government acquisitions of Big Data. Though such a holding would leave unresolved many important questions, it would be a significant first step that would bring the Fourth Amendment into the twenty-first century and enable the next generation of Americans to conduct their lives without fear of unreasonable government searches and seizures of their data.*

* Brad Turner is a graduate of Duke Law School and a practicing attorney in Ohio. I would like to thank the terrific staff of the North Carolina Journal of Law and Technology for their significant contributions to this Article. I would also like to thank my wonderful partner, Kristi Horvath, for her love, support, and frequent reminders that formal writing does not mean "sterile" writing.

TABLE OF CONTENTS

I. INTRODUCTION

"Big Brother is Watching You."[1]

Big Data[2] is fast becoming big business.[3] In an effort to target consumers with advertisements that connect consumers with goods and services that they are likely to buy, businesses track, collect, store, analyze, and share consumer data.[4] From smartphones[5] to smart thermostats,[6] from customer loyalty cards[7] to in-store motion detectors,[8] from cookies[9] to web beacons[10] and beyond,[11]

---

[1] GEORGE ORWELL, 1984, at 2 (Signet Classic 1977).

[2] I use the term "Big Data" throughout this Article to refer to the entire ecosystem of data trackers, collectors, analyzers, sharers, and sellers of people's data. I define data to include any information in electronic form.

[3] *See* McKinsey & Company, *Big Data, Analytics and the Future of Marketing and Sales*, FORBES (July 22, 2013, 9:13 AM), http://www.forbes.com/sites/mckinsey/2013/07/22/big-data-analytics-and-the-future-of-marketing-sales/ (discussing how Big Data is becoming big business).

[4] *See* Lisa Arthur, *What Is Big Data?*, FORBES (Aug. 15, 2013, 8:17 AM), http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/ (discussing how unstructured and structured data collectively makes up "Big Data").

[5] *How Big Data Is Transforming The Mobile Industry*, BUSINESS INSIDER (June 20, 2013, 7:00 PM), http://www.businessinsider.com/big-data-will-transform-mobile-industry-2013-6 (showing how smartphones are part of Big Data collection).

[6] Parmy Olson, *Nest Gives Google Its Next Big Data Play: Energy*, FORBES (Jan. 13, 2014, 6:28 PM), http://www.forbes.com/sites/parmyolson/2014/01/13/nest-gives-google-its-next-big-data-play-energy/ (discussing Google's "next Big Data play:" Nest—a smart thermostat company).

[7] Tom Groenfeldt, *Sears Competes on Big Data and Loyalty Programs*, FORBES (May 2, 2012, 10:20 PM), http://www.forbes.com/sites/tomgroenfeldt/2012/05/02/sears-competes-on-big-data-and-loyalty-programs/ (explaining Sears's Loyalty Card Big Data program).

[8] Paul Richards, *Consumer Behavior Aggressively Tracked This Season*, CNBC (Nov. 29, 2013, 7:34 AM), http://www.cnbc.com/id/101235143 (detailing new technologies businesses have deployed to track consumer behavior).

[9] *See infra* Part II.A.

[10] *Consumers, Big Data, and Online Tracking in the Retail Industry: A Case Study of Walmart*, THE CENTER FOR MEDIA JUSTICE 22 (Nov. 2013), http://centerformediajustice.org/wp-content/files/WALMART_PRIVACY_.pdf (explaining that Walmart uses web beacons to track its users' activities).

companies obtain people's data from every source possible in an attempt to transform that data into consumer sales.[12]

Of course, Big Data is valuable to more than just the private sector. Governments of all shapes and sizes are quickly learning the potential value of obtaining and using Big Data.[13] Unlike business, however, government cannot obtain huge troves of data about its citizens without raising the specter of Orwell's Big Brother.[14] Recent revelations about the size and scope of the National Security Agency's ("NSA") data-collection efforts, for example, have sparked a national debate about the propriety of the government collecting huge quantities of highly-detailed data about its citizens.[15] In a post-9/11 age when people conduct much of their daily lives online, Americans are understandably concerned about whether our national security apparatus strikes the proper balance between national security and civil liberties.[16]

Thankfully, unlike the citizens in Orwell's *1984*,[17] Americans have a tough, old friend to protect them from Big Brother: the

---

[11] *See* Arthur, *supra* note 4 (introducing Big Data and the ways it can be collected).

[12] McKinsey & Company, *supra* note 3 (discussing how Big Data is becoming big business).

[13] *See* Gil Press, *A New Big Data Roadmap for Government (and Business)*, FORBES (Oct. 3, 2012, 6:07 PM), http://www.forbes.com/sites/gilpress/2012/10/03/a-new-big-data-roadmap-for-government-and-business/ (discussing several reports that show governments large and small are learning to embrace Big Data).

[14] *See, e.g.*, Fareed Zakaria, *Big Data, Meet Big Brother*, TIME (July 8, 2013) http://content.time.com/time/magazine/article/0,9171,2146453,00.html (raising the specter of big brother). As an aside, I chose the title of this Article long before finding this similarly-titled article. I take the serendipitous coincidence as reassurance that I am not alone in my concerns and that this Article raises legitimate concerns worthy of public debate.

[15] *See* Eileen Sullivan & Lara Jakes, *NSA Surveillance Debate Muddied by Numbers Game*, HUFFINGTON POST (June 20, 2013, 6:23 PM), http://www.huffingtonpost.com/2013/06/20/nsa-surveillance-debate_n_3475028.html (discussing public debate over the NSA program).

[16] *See id.* (showing how the debate is about proper balance of national security and civil liberties).

[17] GEORGE ORWELL, 1984 (Signet Classic 1977).

Fourth Amendment.[18] The Fourth Amendment's protections against unreasonable searches and seizures have protected Americans' persons, papers, and effects for generations. And in an age when people's lives are constantly being tracked, recorded, analyzed, and shared by third parties,[19] its protections have never been more important.

The problem, exposed by the NSA's continued snooping, is a bit of Fourth Amendment jurisprudence called the "Third-Party Doctrine."[20] Long ago, the Supreme Court said that a person has no reasonable expectation of privacy in information that person knowingly exposes to others.[21] While such a policy may have made sense at a time when ubiquitous government surveillance was a practical and political impossibility, it makes little sense today. Nearly everything people do today becomes data. And nearly every bit of data is shared, knowingly or unknowingly, voluntarily or involuntarily, with others. The script has flipped: it is as difficult today for a person to avoid being tracked as it was thirty or forty years ago for the government to track that same person. Thus, a once small and manageable exception to the Fourth Amendment, the Third-Party Doctrine, now threatens to swallow whole the privacy guaranteed by the Fourth Amendment.

As the institution that created the Third-Party Doctrine so many years ago,[22] courts have the duty to ensure that it does not completely destroy privacy in the information age—an era where constant and pervasive surveillance is the norm. This Article suggests that courts adopt the *Klayman v. Obama*[23] approach and hold that the Fourth Amendment applies to government

---

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

[19] *See* Daniel Zwerdling & G.W. Schulz, *Your Digital Trail, And How It Can Be Used Against You*, NPR (Sept. 30, 2013, 11:00 AM), http://www.npr.org/blogs/alltechconsidered/2013/09/30/226835934/your-digital-trail-and-how-it-can-be-used-against-you (discussing how people are leaving a "digital trail" everywhere they go and with everything they do).

[20] *See infra* Part IV.C (explaining third party rule).

[21] *See infra* Part IV.C.

[22] *See infra* Part IV.C.

[23] 957 F. Supp. 2d 1 (D.D.C. 2013).

acquisitions of Big Data, including metadata. More specifically, courts should follow Justice Alito's reasoning in *United States v. Jones*[24] to hold that government acquisitions of Big Data are searches subject to the reasonableness requirements of the Fourth Amendment. Surely, if the government's collection of someone's global positioning system ("GPS") data in *Jones* was intrusive enough to constitute a search, then so are government acquisitions of Big Data.

Though such a holding would leave unresolved many challenging questions, such as whether the collection of bulk data would require a warrant, it would be an important first step that would bring the Fourth Amendment into the twenty-first century and enable the next generation of Americans to conduct their lives without fear of unreasonable government searches and seizures of their data.

II. BIG DATA AND THE ILLUSION OF PRIVACY

Despite whatever Americans may believe or desire about privacy, their activities are far from private. Every website visit, every hyperlink click, every Facebook message sent, and every YouTube video watched is being tracked.[25] Even offline activities, like shopping, driving, walking, and exercising are being tracked.[26] In the words of the *Jurassic Park* ranger tracked by a pack of velociraptors,[27] "We are being hunted."[28] Instead of velociraptors hunting Americans for lunch, Americans are being hunted for the purpose of advertising, or more generally, for the purpose of making money from their data. Only, unlike the trained *Jurassic Park* ranger, many Americans do not know they are being hunted.

---

[24] 132 S. Ct. 945 (2012).

[25] *See infra* Part II.A.

[26] *See infra* Part II.B.

[27] These small but fast carnivorous dinosaurs make for a surprisingly good metaphor. Translated literally, the term velociraptor means "swift seizer." *Velociraptor*, RED ORBIT, http://www.redorbit.com/education/reference_library/animal_kingdom/dinosauria/1112832683/velociraptor/ (last visited Nov. 21, 2014).

[28] JURASSIC PARK (Universal Pictures 1993).

One scholar compared this to a two-way mirror: the end-user sees her own activities reflected in the two-way mirror, and does not realize that on the other side, she is actually being observed by any number of faceless, non-descript organizations that she probably does not even know exist.[29]

A. *Tracking Online Activities*

Nearly everything a person does online is tracked in some way. Advertisers are looking for valuable ad space, and companies with that space are eager to cash-in.

1. *The Fundamentals of Online Tracking: Cookies, Website Activity Logs, Form Data, and Web Beacons*

When a person visits a website, the website will place a "cookie" on the person's computer or electronic device that tracks the person's activities on the website.[30] These cookies can be set to erase themselves after the individual leaves the website—or not.[31] Persistent (multi-session) cookies stay on a person's computer and can stay there until they expire, which can be months or even years.[32] Persistent cookies are the objects of code on a person's computer that enable a website to "remember" a visitor so that a visitor need not, for example, re-enter her username and password each time she wishes to log on to her email.[33] Typically, cookies

---

[29] *See* Anne Klinefelter, *When to Research Is to Reveal: The Growing Threat to Attorney and Client Confidentiality from Online Tracking*, 16 VA. J.L. & TECH. 1, 4 (2011) (introducing the concept of the "two-way mirror" for the Internet data context). Users who have downloaded and installed "Do Not Track" software, like "Do Not Track Me," will see that for every website they visit, numerous vaguely-named shadowy-sounding organizations like "Qualaroo" and "Typekit" and "Comscore Beacon" are tracking their activities. *See*, *e.g.*, ABINE, https://www.abine.com/donottrackme.html (last visited May 24, 2014) (showing the names of several blocked trackers).

[30] *Internet Cookies*, FED. TRADE COMM'N, http://www.ftc.gov/ftc/cookies.shtm (last visited May 24, 2014); *Internet Explorer 10 Privacy Statement for Windows 7 Last update: December 2012*, MICROSOFT, http://windows.microsoft.com/en-us/Internet-explorer/ie10-win7-privacy-statement (last visited May 24, 2014).

[31] *Internet Cookies*, *supra* note 30.

[32] *Id.*

[33] *Id.*

can be read only by the website generating the cookie.[34] That, however, is not always the case. Sometimes the cookie can be read by multiple websites, permitting information-sharing between several different websites visited by the same person.[35] Thankfully, cookies can be blocked,[36] although often at the cost of severely reduced functionality.[37] New cookie-like technology dubbed "canvas fingerprinting," however, may be "virtually impossible to block."[38]

Cookies are not the only way an online entity can track a user's activities. Users are tracked in any number of ways that do not require the ability to store a cookie on a user's computer or electronic device. For example, websites[39] keep detailed activity logs of every visitor,[40] like an automatically generated visitor log or guestbook. These logs gather raw user-data, like the accessing-device's IP address, the access date and time, and cookie data, if it exists.[41] Software then reads and interprets this raw data to provide the website operator with information about user behavior.[42] Of course, any data directly entered by visitors into

---

[34] *Id.*

[35] *Id.*

[36] *See, e.g.*, *Block or allow cookies*, MICROSOFT, http://windows.microsoft.com/en-us/windows-vista/block-or-allow-cookies (last visited Aug. 9, 2014).

[37] *See, e.g.*, *Cookies Policy*, FUTURE PLC, http://www.futureplc.com/cookies-policy/ (last visited Sept. 20, 2014) (providing options to opt-out of the use of cookies, but at the cost of reduced functionality).

[38] Julia Angwin, *Meet the Online Tracking Device That is Virtually Impossible to Block*, PROPUBLICA (July 21, 2014), http://www.propublica.org/article/meet-the-online-tracking-device-that-is-virtually-impossible-to-block; *see generally* Gunes Acar et al., *The Web Never Forgets: Persistent Tracking Mechanisms in the Wild*, https://securehomes.esat.kuleuven.be/~gacar/persistent/index.html (last visited Sept. 28, 2014) (describing how canvas fingerprinting and other hard-to-block tracking mechanisms work).

[39] Or, more accurately, the server hosting the website.

[40] *See* Sumit Sukhwani, Satish Garla, & Goutam Chakraborty, *Paper 100-2012: Analysis of Clickstream Data Using SAS*, SAS GLOBAL FORUM 2012, 1, 1–2 http://support.sas.com/resources/papers/proceedings12/100-2012.pdf.

[41] *See id.*

[42] *See, e.g.*, SPLUNK, http://www.splunk.com/view/splunk/SP-CAAAG57 (last visited May 24, 2014) (selling data analytics services).

form fields, like name, contact information, etc., is stored in the webserver's database. Cookie data, website activity log data, and form data can then be combined and associated to build a comprehensive snapshot of a particular visitor's activity on the website.[43]

Web beacons are another popular, very simple, and very effective way of tracking people's online whereabouts without using a cookie. Each time someone visits a website embedded with a tracker's web beacon, the tracker is notified, like a blip on a radar.[44] From the blips, the tracker can surreptitiously track a visitor's online whereabouts across any website embedded with the beacon. Embed enough web beacons into enough websites, and a tracker can learn a great deal about a visitor: everything from the visitor's political ideology to the visitor's sexual preferences.

2. *The Incentive to Track and Collect as Much Data as Possible—Big Data's Raison d'Être*

Hosting websites and providing services is not cheap, and it certainly is not free. To the pay the bills, companies often sell ad space to online advertising agencies eager to reach a broader audience.[45] And when selling ad space pays the bills, information about the company's electronic visitors is very valuable. The more information that advertising agency has about a particular company's electronic visitors, the better the ad agency is able to

[43] Much of this is from my own knowledge as a semi-professional web designer with about nine years of experience. But do not take my word for it. One Buzzfeed.com staffer listed eighty things Facebook knows about its users just from form-field data. The list, which is not comprehensive, includes everything from political and religious views to those friends users search for most. Charlie Warzel, *Here Are 80 Deeply Personal Things that Facebook Knows About You Right Now*, BUZZFEED (May 23, 2014, 2:17 PM), http://www.buzzfeed.com/charliewarzel/here-are-80-deeply-personal-things-that-facebook-knows-about.

[44] *See* Joanna Geary, *Tracking the Trackers: What are Cookies? An introduction to web tracking*, THE GUARDIAN (Apr. 23, 2014, 12:08 AM), http://www.theguardian.com/technology/2012/apr/23/cookies-and-web-tracking-intro (explaining web beacons).

[45] *See, e.g.*, GOOGLE ADSENSE, http://www.google.com/adsense (last visited May 24, 2014).

display ads that influence those visitors.[46] The more effective the ad, the more money the advertisement agency makes from its clients.[47] The more money advertising agencies make from advertising on a particular electronic space, the more money the host-company can charge for that ad space.[48]

Such an information and data-driven system encourages ad space hosts, as well as the advertising agencies that buy ad space, to collect as much personal data as possible from visitors. Though a particular website may enjoy access to only its visitors' information, an advertising agency has access to the visitor information of all of the electronic spaces where it displays ads (e.g., websites, smartphone apps, software, console gaming systems, etc.).[49] The ability to compile data from so many different sources helps advertising agencies create a three-dimensional, high-detailed image of any particular visitor.[50] Complex algorithms

---

[46] Jami Oetting, *Advertising's Big Data Debate: 10 Views on Whether Data is Helping or Hurting the Industry*, AGENCY POST (Oct. 22, 2013), http://www.agencypost.com/advertisings-big-data-debate-10-views-on-whether-data-is-helping-or-hurting-the-industry/ (debating the pros and cons of Big Data in the field of advertising). I should also note that generating "impressions" is another way for advertisers to generate revenue, though less than click-throughs. An "impression" is a view. A click-through is, well, a click.

[47] *See How does Google make money? What is driving Google's growth?*, GOOGLE, http://investor.google.com/corporate/faq.html#toc-money (last visited Nov. 14, 2014) ("Our proprietary technology automatically matches ads to the content of the page on which they appear, and advertisers pay us either when a user clicks on one of its ads or based on the number of times their ads appear on the Google Network.").

[48] *See, e.g.*, Sarah E. Needleman & Jack Marshall, *Facebook Ads Become 'Costlier' Choice for Small Businesses*, WALL ST. J. (Aug. 6, 2014, 3:31 PM), http://online.wsj.com/articles/facebook-ads-become-costlier-choice-for-small-businesses-1407341983 (reporting that Facebook is raising advertising rates as it claims its advertising is producing "better outcomes for its advertisers").

[49] *See, e.g.*, *Benefits of Pay-Per-Click (PPC) Advertising*, GOOGLE ADWORDS, http://www.google.com/adwords/benefits/#subid=us-en-et-nelson_adshp_lrn_awhp (last visited Oct. 17, 2014) (touting Google's omnipresence).

[50] *See* Danny Dover, *The Evil Side of Google? Exploring Google's User Data Collection*, THE MOZ BLOG (June 24, 2008), http://moz.com/blog/the-evil-side-of-google-exploring-googles-user-data-collection (documenting all the data Google has admitted to collecting about it users).

then track that visitor's electronic movements and place advertisements on websites, between songs, and before and during videos that are more likely to influence the visitor.[51]

Facebook and Google have become two of the most pervasive data-aggregating advertising agencies. Google's AdSense advertisement system dwarfs most other online ad agencies, boasting more than two million affiliates.[52] Facebook's advertising program is different, but just as massive. Facebook enjoys access to the data of more than 950 million users.[53] Facebook records more than 2.5 billion status updates, wall posts, photos, videos, and comments every day.[54] Facebook also collects data about users visiting any platform that contains a Facebook "Like" button, whether the Facebook user clicks the "Like" button or not.[55] One company estimates that these "Like" buttons exist on nearly one million websites.[56] With access to so much high-quality data, it is no wonder why Google and Facebook are able to target users and visitors with such eerily accurate web ads.[57]

B. *"Offline" Tracking*

Going offline will not stop the hunt; people are tracked even when they think they are offline. Smartphones, GPS systems,

---

[51] *See*, *e.g.*, *About Adsense for Your Blog*, GOOGLE, http://support.google.com/blogger/bin/answer.py?hl=en&answer=42534 (last visited May 24, 2014) (discussing how Google AdSense works).

[52] GOOGLE ADSENSE, *supra* note 45.

[53] Eliza Kern, *Facebook Is Collecting Your Data—500 Terabytes a Day*, GIGAOM (Aug. 22, 2012, 3:25 PM), http://gigaom.com/2012/08/22/facebook-is-collecting-your-data-500-terabytes-a-day/.

[54] *Id.* Facebook receives more than 2.7 billion "Likes" per day. *Id.* Altogether, more than 500 terabytes (500,000 gigabytes or 0.5 petabytes) of new data is recorded by Facebook alone every single day. *Id.* Just one of Facebook's largest storage server clusters hosts more than 100 petabytes of data. *Id.*

[55] Riva Richmond, *As 'Like' Buttons Spread, So Do Facebook's Tentacles*, N.Y. TIMES (Sept. 27, 2011, 3:51 PM), http://bits.blogs.nytimes.com/2011/09/27/as-like-buttons-spread-so-do-facebooks-tentacles/.

[56] *Id.*

[57] *See, e.g.*, Brad Stone, *Ads Posted on Facebook Strike Some as Off-Key*, N.Y. TIMES (Mar. 3, 2010), http://www.nytimes.com/2010/03/04/technology/04facebook.html?_r=0.

customer-loyalty cards, video cameras, and even radio frequency identification ("RFID") devices track the whereabouts, purchasing habits, and seemingly offline activities of Americans every single day, even though Americans may not realize it or approve.

1. *Smartphones and Things that Connect to Smartphones*

Smartphones can do much more than browse the Internet and host apps. Modern smartphones come equipped with microphones, high-quality cameras, GPS chips, compasses,[58] accelerometers,[59] and gyroscopes.[60] Separately, each of these on-board devices can collect a unique type of personal data that in itself can be highly revealing. But together, these devices can create an intimately detailed picture of an individual's personal activities.

This functionality is so revolutionary that it has given birth to an entire movement, called "The Quantified Self" ("QS"). According to Cisco, a technology company apparently hoping to capitalize on the QS movement, the movement "employs technology to drive greater self-awareness by tracking data related to exercise, diet, health maintenance, financial management, learning, and so forth."[61] The website quantifiedself.com lists more than 500 smartphone apps capable of helping people track themselves.[62] For example, one app called "Fitbit" tracks users' physical activity and sleep.[63] Once recorded, user data is uploaded

[58] Sara M. Watson, *The Latest Smartphones Could Turn Us All into Activity Trackers*, WIRED (Oct. 10, 2013, 9:29 AM), http://www.wired.com/opinion/2013/10/the-trojan-horse-of-the-latest-iphone-with-the-m7-coprocessor-we-all-become-qs-activity-trackers/.

[59] *Id.* An accelerometer measures the proper acceleration of an object. *Accelerometer*, THE FREE DICTIONARY, http://encyclopedia2.thefreedictionary.com/Accelerometer (last visited Nov. 21, 2014).

[60] *See* Watson, *supra* note 58. A gyroscope is used to detect the precise orientation of an object in three-dimensional space. *Gyroscope*, DICTIONARY.COM, http://dictionary.reference.com/browse/Gyroscope (last visited Nov. 21, 2014).

[61] Joseph Bradley, *When IoE Gets Personal: The Quantified Self Movement!*, CISCO BLOG (Sept. 10, 2013, 1:33 PM), http://blogs.cisco.com/zzfeatured/when-ioe-gets-personal-the-quantified-self-movement/.

[62] QUANTIFIED SELF, http://quantifiedself.com/guide/ (last visited May 25, 2014) (listing over 500 tools).

[63] *Id.*

to Fitbit, which then processes the data and provides users with visualizations of their physical activity and sleep.[64] Another app, "Digifit," is a "full suite of Apple apps that records heart rate, pace, speed, cadence, and power of your running, cycling and other athletic endeavors."[65] Of course, once collected, the data can then be uploaded to various athletic-training websites.[66] "Moodpanda" tracks a user's mood.[67] "Momento" allows users to type journals entries, upload photos, and then tag them with contacts in the user's address book, GPS location data, and other tags.[68]

As revolutionary as it is already, the QS movement has just begun. In addition to the suite of chips and sensors onboard smartphones, the increasing availability of cheap sensors outside of the smartphone is driving the QS movement even smaller, even more precise, and even more pervasive.[69] Soon, clothing,[70] shoes,[71] headbands,[72] and pills[73] will be embedded with small sensors[74] to help smartphones track every moment, every breath, every blink, every sugar high, every sugar low, every REM cycle, and, yes, every bowel movement.

The QS movement, as invasive as it is, is at least the result of the people's choice to track themselves. Businesses, however, do not always inform customers of their efforts to track consumer data. Google has been caught secretly tracking users' walking and

---

[64] *Id.*
[65] *Id.*
[66] *Id.*
[67] *Id.*
[68] *Id.*
[69] Joseph Bradley, *A Pill That Tracks Your Health? The Reality Of the 'Quantified Self' Movement*, HUFFINGTON POST (Nov. 18, 2013, 5:12 AM), http://www.huffingtonpost.com/joseph-bradley/cisco-quantified-self-movement_b_3907545.html.
[70] *Id.*
[71] *Id.*
[72] *Id.*
[73] *Id.*
[74] *Id.*

bicycling activity.[75] Apple, through its iPhone, was caught tracking users' geographic locations based upon cellphone tower and Wi-Fi-network triangulation.[76] One company in London, called "Renew," is beta-testing the use of Wi-Fi-enabled trash cans to collect information from passersby's smartphones, without any warning whatsoever.[77] Most recently, Microsoft was caught reading at least one of its user's emails just after launching a public advertisement campaign against Google for similar practices.[78]

2. *Cameras, Loyalty Cards, and In-Store Tracking*

Smartphones are not the only way businesses can watch and learn from their customers. Even customers who do not have a computer of any kind cannot escape the ever-watchful eyes of businesses seeking to learn more about their customers.

Though businesses are reluctant to talk about it, some businesses, especially big-chain stores, begin monitoring a customer the moment the customer parks in the store's parking

[75] Kashmir Hill, *Big Healthy Brother: 'Google Now' Surprises Users by Tracking Miles Walked and Biked*, FORBES (Nov. 2, 2012, 12:18 PM), http://www.forbes.com/sites/kashmirhill/2012/11/02/big-healthy-brother-google-now-surprises-users-by-tracking-miles-walked-and-biked/.

[76] Watson, *supra* note 58.

[77] Anthony Gucciardi, *Wi-fi Trashcans Now Silently Tracking Your Smartphone Data*, STORY LEAK (Aug. 12, 2013, 3:15 AM), http://www.storyleak.com/wi-fi-trashcans-tracking-your-smartphone-data/. Though this occurred in London, there is no reason to think that this couldn't happen in the United States, especially given that Europe protects people's privacy much more rigorously than the United States. *See, e.g.*, Daniel Fisher, *Europe's 'Right To Be Forgotten' Clashes With U.S. Right To Know*, FORBES (May 16, 2014, 8:45 AM), http://www.forbes.com/sites/danielfisher/2014/05/16/europes-right-to-be-forgotten-clashes-with-u-s-right-to-know/ (reporting on how Europe's robust effort to protect personal privacy, but noting that this may just reflect a difference in Europe's philosophy about the role of government in protecting people's privacy).

[78] Javed Anwer, *Caught Red-handed, Microsoft Promises Not to Snoop on Emails*, TIMES OF INDIA (Mar. 29, 2014, 3:03 AM), http://timesofindia.indiatimes.com/tech/tech-news/Caught-red-handed-Microsoft-promises-not-to-snoop-on-emails/articleshow/32870284.cms.

lot.[79] Video cameras around the parking lot record the vehicle and license plate.[80] Once the customer enters the store, cameras hidden in tiny holes in shelving, the eyes of mannequins, and elsewhere record the customer's movement, age, sex, ethnicity, and facial expressions as she moves around the store and interacts with various products for sale.[81] When the customer purchases something with a credit card or uses a customer-loyalty card, the store then associates the in-store behavioral data with a real person and that person's account information.[82]

Google recently released[83] a new analytics program called "Universal Analytics" that tracks a person's physical location all of the time—even when the person is not using any Google apps—to enable big business to tie a person's out-of-store data with a person's in-store behavior.[84] Google has been tracking people's GPS location for years,[85] but unlike regular GPS-tracking, Universal Analytics detects when a person enters a store and then

---

[79] *See* Chris Moran, *4 Ways Retail Stores Are Monitoring Your Every Move*, CONSUMERIST (Mar. 27, 2013), http://consumerist.com/2013/03/27/4-ways-retail-stores-are-monitoring-your-every-move/.

[80] *Id.*

[81] *Id.* These tactics may seem extreme, but businesses believe that these tactics help them to truly understand their customers. For example, video cameras in mannequins track eye movement and facial expressions because businesses want to know whether a consumer reacts favorably or unfavorably to certain products. Businesses stick video cameras in shelves to see whether consumers read labels, search for deals, or simply snap-up their favorite items. The additional information about age, sex, and ethnicity helps businesses to learn about whether different demographic groups react differently to the same product. Knowing, for example, that women select a particular food product more than men might help that retailer more efficiently target its ads.

[82] *See id.* (discussing how the data can be tied to other data).

[83] Kristian Petterson, *Google's Universal Analytics – Ready to Jump In?*, BUSINESS2COMMUNITY (Oct. 24, 2014), http://www.business2community.com/business-intelligence/googles-universal-analytics-ready-jump-01046094.

[84] John McDermott, *Google Takes Its Tracking into the Real World*, DIGIDAY (Nov. 6, 2013), http://digiday.com/platforms/google-tracking/; Paul Lear, *In-Store (Offline) Tracking with Google's Universal Analytics*, BLAST AM (Aug. 25, 2013), http://www.blastam.com/blog/index.php/2013/08/offline-tracking-with-universal-analytics/.

[85] McDermott, *supra* note 84.

informs the store of the user's Google searches for that business.[86] One author described this kind of offline-online data tracking as the "Holy Grail" of data analytics.[87]

3. *The "Internet of Things"*

The "Internet of Things" refers to the prospect that nearly everything that can be connected to the Internet will be in the near future.[88] According to one study, by the year 2020, more than 30 billion devices could be wirelessly connected to the Internet.[89] Everything from televisions to refrigerators to electricity meters will be capable of recording data and transferring that data to third parties, with or without a user's knowledge or consent.

Of course, the "Internet of Things" may provide significant benefits, such as a refrigerators that can warn people that food has gone bad,[90] or smart electricity meters that permit people to make more informed decisions about their electricity usage.[91] But the data-collection and data-sharing activities of things connected to the Internet is not always so obvious to consumers. Nor is it always a choice. Take for example, the story of the British blogger, Doctorbeet.[92] After purchasing a new LG Smart TV, he discovered that it was secretly logging his viewing activities and uploading the

---

[86] *Id.*

[87] Petterson, *supra* note 83.

[88] *See* Catherine Crump & Matthew Harwood, *Big Brother Is Coming: Google, Mass Surveillance, and the Rise of the "Internet of Things*,*"* SALON (Mar. 26, 2014, 8:59 AM), http://www.salon.com/2014/03/26/big_brother_is_here_google_mass_surveillance_and_the_rise_of_the_Internet_of_things_partner/ (defining the "Internet of Things"). Perhaps the biggest supporter of the Internet of Things is Cisco Systems, one of the world's largest manufacturers of networking equipment. Cisco calls it the "Internet of Everything," or IoE. *The Internet of Everything*, CISCO, http://www.cisco.com/web/about/ac79/innov/IoE.html (last visited May 25, 2014).

[89] *See* Crump & Harwood, *supra* note 88.

[90] *Id.*

[91] *See*, *e.g.*, *What Is A SmartMeter?*, PG&E, http://www.pge.com/en/myhome/customerservice/smartmeter/index.page (last visited Oct. 17, 2014) (discussing the benefits of smart meters).

[92] *See* Crump & Harwood, *supra* note 88 (describing Doctorbeet's experience with a snooping television).

data, unencrypted, to LG.[93] Doctorbeet then discovered that this was a default feature.[94] But even after turning off the feature, the television continued to monitor and share his viewing activity.[95]

Or consider that, starting in September 2014, a recent federal government regulation will require manufacturers to equip nearly every vehicle they produce with a "black box" containing an "Event Data Recorder."[96] The data recorder will constantly monitor the host-vehicle's speed, braking, driving patterns, and yes, location.[97] It is unclear who owns the recorder and the data it collects.[98] At this point, there is nothing to stop insurance companies, for example, from contractually requiring policyholders to submit the black boxes to regular inspection, storage, and analysis, whether or not the vehicle's owner also owns the black box and the data it collects. Subprime auto-title lenders already track the GPS location of the vehicles debtors pledge as collateral.[99] These lenders even have the ability to remotely prevent a debtor's vehicle from starting when that debtor misses a payment.[100]

C. *Big Data*

Data collected by private parties, online and off, is often sold to additional private parties. In fact, selling databases of information about user activity has become a multi-billion dollar industry of

---

[93] *Id.*
[94] *Id.*
[95] *Id.*
[96] Advertorial Published in Wheels, *New Laws Emerge as Technology Impacts Cars and Driving*, PITTSBURGH POST-GAZETTE (Mar. 27, 2014, 5:00 AM), http://www.post-gazette.com/business/auto/2014/03/27/New-laws-emerge-as-technology-impacts-cars-and-driving/stories/201403270171.
[97] *Id.*
[98] *Id.*
[99] *See* Robert Szypko, *Your Car Won't Start. Did You Make The Loan Payment?*, NPR (Oct. 16, 2014, 2:21 PM), http://www.npr.org/blogs/alltechconsidered/2014/10/16/356693782/your-car-wont-start-did-you-make-the-loan-payment (reporting about data-collection and repossession activities of subprime auto-title lenders).
[100] *See id.*

trackers, aggregators, analyzers, and buyers of people's data[101] known collectively as "Big Data."[102] The companies that collect user data and then resell it as part of the Big Data industry are often called "data brokers."[103] The largest Big Data broker is Acxiom.[104] Acxiom claims that its database contains about 1,500 data points for each of the 500 million people it tracks.[105]

The reach of data brokers is so vast that it recently drew the attention of the United States Senate.[106] The Senate Commerce Committee investigated the nine largest data aggregators' "collection, use, and sale of consumer data for marketing purposes."[107] The committee learned that Big Data brokers collect a "huge volume of detailed information on hundreds of millions of consumers," which includes offline data.[108] The committee report also noted that these brokers "amass data without the direct interaction with consumers" and even contractually prohibit the businesses that purchase data from these brokers from disclosing to

---

[101] *See* Dave Feinleib, *The Big Data Landscape*, FORBES (June 19, 2012, 1:01 PM), http://www.forbes.com/sites/davefeinleib/2012/06/19/the-big-data-landscape/ (discussing the Big Data ecosystem).

[102] *See* Steve Lohr, *IDC Sizes Up the Big Data Market*, N.Y. TIMES BITS BLOG (Mar. 7, 2012, 12:51 PM), http://bits.blogs.nytimes.com/2012/03/07/idc-sizes-up-the-big-data-market/?_php=true&_type=blogs&_r=0 (discussing the "Big Data" industry).

[103] Steve Kroft, *The Data Brokers: Selling Your Personal Information*, CBS NEWS (Mar. 9, 2014), http://www.cbsnews.com/news/the-data-brokers-selling-your-personal-information/.

[104] *Id.*

[105] Natasha Singer, *Mapping, and Sharing, the Consumer Genome*, N.Y. TIMES (June 16, 2012), http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?pagewanted=all.

[106] *See generally* OFFICE OF OVERSIGHT & INVESTIGATIONS MAJORITY STAFF, S. COMM. ON COMMERCE, SCI., AND TRANS., 113TH CONG., 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=bd5dad8b-a9e8-4fe9-a2a7-b17f4798ee5a (investigating the activities of data aggregators).

[107] *Id.*

[108] *Id.*

consumers the sources of consumer data—something the committee called a "veil of secrecy."[109]

Facebook and Google face continuing questions about what they do with all of the data they collect. For example, one Bulgarian blogger and digital-rights activist allegedly purchased over 1.1 million Facebook data-entries, which included information like Facebook names, user IDs, and email addresses.[110] A Facebook investigation found that the data had been obtained by merely "scrapping" (harvesting with bots) users' "public" data.[111] But some question this conclusion because at least some of the users' email addresses were not publically available.[112] And even if Facebook does not "sell" user data, it still permits other parties to access user data.[113] Application developers that create apps for Facebook need only receive a Facebook user's click of the "I agree" button before the user's data is sent directly to the developer.[114] What those developers do with the data is unknown, though Facebook's contract with developers states, "You will not sell user data."[115]

D. *Big Data Will Know You Better Than You Know Yourself*

Data's value is in its predictive powers. There is little benefit to recording, storing, sharing, and aggregating data about people if doing so does not accurately describe people's habits, behaviors,

---

[109] *Id.*

[110] Andy Greenberg, *Facebook Investigating How Bulgarian Man Bought 1.1 Million Users' Email Addresses for Five Dollars*, FORBES (Oct. 25, 2012, 4:39 PM), http://www.forbes.com/sites/andygreenberg/2012/10/25/facebook-investigating-how-bulgarian-man-bought-1-1-million-users-email-addresses-for-five-dollars/.

[111] *Id.*

[112] *Id.*

[113] *See* Max Read, *How to Stop Facebook from Sharing Your Information with Third Parties*, GAWKER (Oct. 18, 2010, 1:15 AM), http://gawker.com/5666325/how-to-stop-facebook-from-sharing-your-information-with-third-parties (showing how to stop Facebook from sharing app data with third parties).

[114] Greenberg, *supra* note 110.

[115] *Statement of Rights and Responsibilities*, FACEBOOK, https://www.facebook.com/legal/terms (last visited Oct. 4, 2014).

and preferences.[116] The inability of people to know themselves, or be honest with themselves, is one of the central premises of the QS movement, and why the movement holds great promise for improving people's health.[117]

But individual tidbits of data can be deceptive. Statisticians and sophisticated consumers alike know the value of a good sample size.[118] Small sample sizes are prone to error because in a small sample, no one knows whether the data is anomalous or consistent with the population as a whole.[119] For example, the voting habits of Orange County, California are historically inconsistent with the voting habits of California as a whole.[120] A politician running in Los Angeles County would do well to ignore the voter data from Orange County[121] and visa-versa.

The same can be said of an individual. One bit of data about a particular consumer—she bought batteries—is of little value to advertisers. Add to that other data—she bought batteries from a sex shop—and advertisers learn something potentially intimate and valuable about that consumer. Add even more data—that the sex shop was 1,000 miles from her home and that the woman is 65 years old—and the advertiser may question the implications raised

---

[116] One could make the argument that individuals who value privacy stand to benefit from recording inaccurate data about themselves. A pro-privacy startup might turn this tactic into a profitable business model. Still, it probably would not prevent businesses from tracking and collecting data.

[117] *See* Gary Wolf, *The Data-Driven Life*, N.Y. TIMES (Apr. 28, 2010), http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=all (extolling the virtues of the QS movement).

[118] *See, e.g.*, *Sample Size Determination*, DEFINITIONS, http://www.definitions.net/definition/sample%20size%20determination (last visited Nov. 21, 2014).

[119] *Id.*

[120] *See* Adam Nagourney, *Orange County Is No Longer Nixon Country*, N.Y. TIMES (Aug. 29, 2010), http://www.nytimes.com/2010/08/30/us/politics/30orange.html (discussing Orange County as a historically Republican stronghold in the middle of a state that is known for being a Democratic stronghold).

[121] *See, e.g.*, *County of Los Angeles Registrar-Recorder/County Clerk Election Results*, LOS ANGELES COUNTY REGISTRAR-RECORDER/COUNTY CLERK, http://rrccmain.co.la.ca.us/14062043/2043_GOVERNOR_Frame.htm (last visited June 8, 2014) (showing that, for example, in the last election the county voted overwhelmingly for a Democrat for Governor).

by the smaller set of data. Finally, add in a little bit more data—the woman bought an insulin pump six months before, and the particular device she bought is known for battery trouble—and the data begins to suggest that the woman did not purchase batteries for a sex device. Rather, it would appear that the woman bought batteries at a sex shop because she was diabetic and desperate to replace the dead or dying batteries in her insulin pump. Advertising sex toys to this woman might generate far less interest than advertising medical devices associated or related to diabetes. The advertiser might further benefit from this knowledge by advertising this woman "long-lasting" batteries or "energy-efficient" insulin pumps. This would not only increase profits for the company, it would also provide a helpful service to the woman. The point being, of course, that the more data an advertiser has about an individual person, the more successful an advertiser will be at selling that person goods and services.

The incentive, therefore, is to collect as much data as possible about every single individual, from every possible angle, and from every possible quadrant of their lives. Data collected in-store is more valuable when paired with data collected through the smartphone. That data is even more valuable if paired with data about what happens inside the person's home. And so on and so forth, until the advertiser has a rich trove of data so personal, so intimate, and so private, that the advertiser knows the person better than the people around that person do, maybe even better than the person knows herself.

Google's entry into industries seemingly unrelated to email or search engines makes sense if it is seen not as a tech company, but rather as a Big Data advertising company seeking to dominate the future of advertisement. Google's desire to collect as much information about people as it can from as many sources as possible can then be explained. If Google can combine data from people's email, smartphone, thermostat,[122] robots,[123] contact

[122] Aaron Tilley, *Google Acquires Smart Thermostat Maker Nest for 3.2 Billion*, FORBES (Jan. 13, 2014, 4:18 PM), http://www.forbes.com/sites/aarontilley/2014/01/13/google-acquires-nest-for-3-2-billion/.

lenses,[124] glasses,[125] self-driving car,[126] etc., then it can gain a significant and lasting advantage over its fellow advertisers who may not have such pervasive access to people's lives. Google's Director of Engineering, Ray Kurzweil, all but confirmed that this is Google's goal when he said that he believes that Google will soon "know the answer to your question before you have asked it," and "will have read every email you've ever written, every document, every idle thought you've ever tapped into a search-engine box."[127] He added that "[Google] will know you better than your intimate partner does," and "[b]etter, perhaps, than even yourself."[128]

E. *Data vs. Metadata*

Data can, arguably, be divided into two types: the content of a person's communications on the one hand and the non-content information about the person's communications on the other.[129] The content of a person's communications is the message itself—*the* speech, *the* picture, *the* text, *the* email, etc.[130] Metadata

---

[123] Steve Henn, *With Google's Robot-Buying Binge, A Hat Tip To The Future*, NPR (Mar. 17, 2014, 4:28 PM), http://www.npr.org/blogs/alltechconsidered/2014/03/17/290888529/with-googles-robot-buying-binge-a-hat-tip-to-the-future ("In less than a year, Google has bought more than a half-dozen robotics companies, setting the industry abuzz.").

[124] *Introducing Our Smart Contact Lens Project*, GOOGLE BLOG (Jan. 16, 2014, 4:18 PM), http://googleblog.blogspot.com/2014/01/introducing-our-smart-contact-lens.html.

[125] *Google Glass*, GOOGLE, http://www.google.com/glass/start/ (last visited May 25, 2014).

[126] Chris Urmson, *The Latest Chapter for The Self-Driving Car: Mastering City Street Driving*, GOOGLE BLOG (Apr. 28, 2014), http://googleblog.blogspot.com/2014/04/the-latest-chapter-for-self-driving-car.html.

[127] *Google Will Soon Know You Better than Your Spouse Does*, *Top Exec Says*, HUFFINGTON POST (Feb. 23, 2014, 1:59 PM), http://www.huffingtonpost.com/2014/02/23/ray-kurzweil_n_4842972.html.

[128] *Id.*

[129] *See, e.g.*, *Metadata: Piecing Together a Privacy Solution*, ACLU OF CAL. (Feb. 2014), https://www.aclunc.org/sites/default/files/Metadata%20report% 20FINAL%202%2021%2014%20cover%20%2B%20inside%20for%20web%20%283%29.pdf (introducing the public to metadata).

[130] *See id.* at 3 (defining "content").

is information *about* the communication minus the message itself.[131] Metadata answers questions like, when did the person send the message? To whom? From what account? How? From what geographic location? How many characters was the text? Did the email include a picture? How big was the picture? Did the picture include any information about who created it?[132] And so on.

Those first learning about metadata might mistakenly believe that metadata is too removed from the content of the data to say anything valuable. But this belief about metadata, however understandable, could not be further from the truth.[133] Metadata can reveal as much if not more than the content of the communication itself. Imagine a simple text message that reads, "Here." The content of the communication is vague and reveals almost nothing at all. Someone is somewhere. But what if the metadata revealed that this message was sent by a woman from an abortion clinic to a man at his home? All of the sudden, the picture is clear. In this case, the content of the message conveyed to the man is far less intimate and revealing than the metadata about that message.

It is no wonder why the cover of a new American Civil Liberties Union ("ACLU") report on metadata features the picture of a woman's face layered over pieces of a puzzle.[134] In a way, metadata can be pieces of a puzzle that, when put together, reveal something quite significant. In its report, the ACLU notes the extent of information that can be revealed by metadata.[135] Metadata reveals everything from "our presence at a hospital, a political rally, or a religious ceremony," to "our calls to an addiction support hotline, a job recruiter, or a dating service," and "our purchases of birth control or books on fighting depression."[136] Metadata can, perhaps more than the content of a message, be "inherently communicative."[137] The ACLU explains that "it's not

[131] *See id.* (distinguishing "metadata" from "content").
[132] *See id.* (defining "metadata").
[133] *See id.* at 5 (explaining how metadata can be so revealing).
[134] *Id.* at 1.
[135] *See id.* at 5 (discussing "communicative" metadata).
[136] *Id.*
[137] *Id.*

hard to uncover an individual's dissatisfaction with his job, marital difficulties, or health status if you know that he is in frequent contact with a recruiter, divorce attorney, or cancer treatment center."[138] In another example, the ACLU adds, "it's even easier to 'infer' a person's sexual orientation or political allegiance if you can reveal her connection with the LGBT Choir or the local Tea Party chapter."[139]

The distinction between metadata and the content of data can also be unclear. Consider URLs. URLs are simultaneously content and metadata. The text of the URL, if typed by a person into a web browser's address bar, is the content of a communication. A person literally types letters and numbers into a text box. It is the digital equivalent of an eighteenth-century author quilling a missive to one person requesting information about another. The content of that missive—the request for information—is, without a doubt, content, not metadata. But, at the same time, the text typed is also an address—a destination. It is therefore not only the question quilled into the missive itself, but also the *place* where the courier is to deliver the missive—metadata. Thus, it is content and non-content, both at the same time.

The same is true even when a person does not type a URL directly into her Internet browser's address bar, but rather types a sentence into a search box. The search terms, like the inquiry quilled into the missive, are the content of the communication. The web user is asking Google to find information. On the other hand, when the user clicks the Google search button, the text that appears in her Internet browser's address bar is "https://www.google.com/#q=oklahoma+law+on+abortions."[140] That text is an address, a destination—metadata. But it is all too easy to figure out from that metadata that this person searched for "Oklahoma law on abortions." So, the search is both data and metadata, content and non-content.

---

[138] *Id.*
[139] *Id.*
[140] See for yourself. Google "Oklahoma law on abortions."

The distinction between metadata and the content of data is probably irrelevant to the average person. People, like the woman in the abortion clinic and the man at home, probably care far more about whether their messages are received by the intended recipient and whether they are kept private. The distinction is nonetheless critical because the applicability of the Fourth Amendment to their messages—and its many privacy protections—is currently determined by this seemingly arbitrary distinction.

F. *The Myth of "Anonymous" Data*

Most Internet users would like to be anonymous when online.[141] Indeed, one study showed that "86% percent of Internet users have taken steps online to remove or mask their digital footprints."[142] Companies are aware of people's privacy concerns and often justify their data-collection programs by claiming that the data they collect is anonymous and that it does not reveal any personally-identifiable information.[143] Many people then use those companies' services under the mistaken belief that their activities, including their online and offline activities and whereabouts, are anonymous and will not, at some later date, come back to haunt them.

But electronic anonymity is a myth.[144] People's online, and

---

[141] *See* Lee Rainie et al., *Pew Internet, Anonymity, Privacy, and Security Online*, PEW RESEARCH INTERNET PROJECT (Sept. 5, 2013), http://pewInternet.org/Reports/2013/Anonymity-online.aspx (polling on online anonymity).

[142] *Id.*

[143] *See*, *e.g.*, *Collection of Anonymous Location Data*, GOOGLE, https://support.google.com/gmm/answer/2839958?hl=en (last visited May 25, 2014) ("You can help improve Google, including products and services like traffic, by allowing location data to be anonymously crowdsourced by Google's location service from your device.").

[144] *See* Kim Zetter, *Anonymous Phone Location Data Not So Anonymous, Researchers Find*, WIRED (Mar. 27, 2013, 4:10 PM), http://www.wired.com/threatlevel/2013/03/anonymous-phone-location-data/ ("Anonymized mobile phone location data produces a GPS fingerprint that can be easily used to identify a user based on little more than tracking the pings a phone makes to cell towers, a new study shows."); *Amazon's Online Workforce Not So Anonymous*

increasingly offline, activities conducted under the belief that those activities were private or at least anonymous, have come back to haunt them. Consider AOL's disastrous but purposeful disclosure of more than 20 million search queries made by over 650 thousand of its users back in 2006.[145] The AOL team that released the data said that the posting was meant to benefit academic research.[146] Once AOL realized what the supposedly anonymized data revealed about their users, it quickly removed the data, but not before the damage had been done.[147] Though the AOL team had attempted to anonymize the data by replacing the AOL usernames with generic numbers, de-anonymizing the data proved all too easy due to the nature of the data collected on each user. The New York Times detailed the relatively simple process reporters used to identify one user.

> No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."
>
> And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia [sic]."
>
> It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga.,

---

After All, UNIV. OF TEXAS AT AUSTIN (Mar. 27, 2013), http://www.utexas.edu/news/2013/03/27/amazons-online-workforce-not-so-anonymous-after-all/ ("Most people assume that Amazon.com's massive online workforce is anonymous, but a study by researchers from The University of Texas at Austin and five other universities has uncovered a security vulnerability that makes it relatively easy to uncover many workers' personally identifying information."); Yves-Alexandre de Montjoye et al., *Unique In The Crowd: The Privacy Bounds Of Human Mobility*, SCIENTIFIC REPORTS (Mar. 25, 2013), http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html ("[I]n a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals.").

[145] Dawn Kawamoto & Elinor Mills, *AOL Apologizes for Release of User Search Data*, CNET (Aug. 7, 2006, 2:30 PM), http://news.cnet.com/AOL-apologizes-for-release-of-user-search-data/2100-1030_3-6102793.html.

[146] *Id.*

[147] *Id.*

> frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.[148]

Other users' searches were not so benign. Consider what AOL user No. 2281868's searches revealed about him or her. According to Consumerist.com, the user is "into jazz . . . bestiality . . . pictures of old ladies who look like Hilary [sic] Clinton naked . . ." and "wants to find black gay overbite porn . . . ."[149] The list continues.[150] Another user's bizarre and somewhat frightening Internet searches inspired the creation of a play called "User 927." The play, a self-described "thriller about cyberstalking, search engines, and the way information is obtained, manipulated, and released in our wired world," received a staged reading in New York in 2009.[151] Privacy advocates still maintain a searchable database of data derived from the AOL disclosure as a warning to people everywhere about how dangerous it is to release such "anonymized" information.[152] It is not hard to see how the release of such data could generate enormous reputational damage.[153] It is

---

[148] Michael Barbaro & Tom Zeller Jr., *Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0.

[149] *AOL User 2281868 Looking for Gay Black Superman with an Overbite*, CONSUMERIST (Aug. 9, 2006), http://consumerist.com/2006/08/09/aol-user-2281868-looking-for-gay-black-superman-with-an-overbite/. Beware, I have only listed some of the user's searches. Some of the user's searches are even more graphic and may be upsetting.

[150] *Id.*

[151] Ben Popken, *AOL User 927 Gets Staged Reading in New York*, CONSUMERIST (June 8, 2009), http://consumerist.com/2009/06/08/aol-user-927-gets-staged-reading-in-new-york/.

[152] *See Search AOL User Searches Like the Pros*, CONSUMERIST (Aug. 9, 2006), http://consumerist.com/2006/08/09/search-aol-user-searches-like-the-pros/ (retaining the search records for all to see). Again, just as a warning, user searches can be graphic and may be upsetting. Proceed only if you are an adult and only if you are capable of maturely handling real people's very real Internet searches.

[153] As harmful to privacy as the AOL disclosure was, the data releases represent only a tiny chunk of the data available on the Internet. Since 2006, databases have grown to incomprehensible size and complexity, connecting user actions to personalized user data. Since 2006, sophisticated computer programs

also not hard to see that these users mistakenly believed their searches were either private or anonymous, and depended on that belief.

III. HOW AND WHY BIG BROTHER OBTAINS BIG DATA

The Third-Party Doctrine holds that the protections of the Fourth Amendment do not apply to information that a person knowingly exposes to others.[154] The doctrine, in theory, renders the Fourth Amendment protections against unreasonable search and seizure almost totally inapplicable to data obtained by the government through private, nongovernmental parties, like businesses.[155] Having detailed the extent to which data is collected by private parties, this section explains why government would want to obtain Big Data and details how the government regularly taps into that data, thus blurring the lines between first-hand government surveillance, which is regulated by the Fourth Amendment, and second-hand surveillance through private parties, which currently is not.[156]

A. *Why Government Wants Big Data*

From national security to healthcare to income eligibility,[157] the government has plenty of reasons to obtain Big Data. According to

---

have been developed to help advertisers understand such immense piles of data, making it even more likely now that a small amount of information would reveal the identity of a user. Moreover, de-anonymizing data requires very little information. Even a small dataset of information associated with real names can be used as a cross reference to de-anonymize a much bigger database. One set of researchers found that "87 percent of the population in the United States, 216 million of 248 million, could likely be uniquely identified by their five-digit ZIP code, combined with their gender and date of birth." Bruce Schneier, *Why 'Anonymous' Data Sometimes Isn't*, WIRED (Dec. 13, 2007), http://archive.wired.com/politics/security/commentary/securitymatters/2007/12/securitymatters_1213.

[154] *See infra* Part IV.C.

[155] *See infra* Part IV.C.

[156] *See infra* Part IV.B.

[157] *See* Melanie Hicken, *What Information Is the Government Buying About You*, CNN (Oct. 30, 2013, 12:35 PM), http://money.cnn.com/2013/10/30/pf/government-data-broker/ (documenting government data collection and why).

its website, IBM's Big Data services can help the government with threat prediction and prevention; social program fraud, waste, and errors; tax compliance, fraud, and abuse; and crime prediction and prevention.[158] IBM adds, "From crime prevention to transportation, defense, national security, revenue management, environmental stewardship and social services, governments must wrestle every day with managing and using this data."[159] According to the McKinsey Global Institute's "Big Data Report," governments in Europe "could save more than €100 billion ($149 billion) in operational efficiency improvements alone by using big data."[160] The estimated amount does not include potential savings from using Big Data to reduce fraud and errors and to boost the collection of tax revenues.[161] And according to a popular tech website for government administrators, Big Data offers administrators the ability to "make better decisions" faster, "improve mission outcomes," "identify and reduce inefficiencies," identify and "eliminate waste, fraud, and abuse," "improve productivity," "boost ROI," and "enhance transparency and service."[162]

Even if these benefits are overblown, it is not hard to see that government, if it has not already, will turn to Big Data in the hope of achieving better outcomes for its citizens. In March of 2012, the White House, along with several other governmental agencies, announced a $200 million dollar "Big Data Research and Development Initiative."[163] The initiative promises to "greatly

---

[158] *Big Data for Government*, IBM, http://www-01.ibm.com/software/data/bigdata/industry-government.html (last visited June 1, 2014).

[159] *Id.*

[160] James Manyika et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, MCKINSEY GLOBAL INSTITUTE, 1 (June 2011), http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.

[161] *Id.*

[162] *8 Benefits of Big Data for State and Local Governments*, STATE TECH MAGAZINE (May 30, 2013), http://www.statetechmagazine.com/article/2013/05/8-benefits-big-data-state-and-local-governments.

[163] Press, *supra* note 13.

improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data."[164] And according to one writer, state and local governments have already tapped into the power of Big Data to "relieve traffic congestion, monitor public utilities, evaluate and predict crime, follow education trends, and keep tabs on public resources."[165]

B. *Distinguishing First-Hand Data Collection from Second-Hand Data Collection*

The Government can gather data in a variety of ways. Those methods can be divided into two basic categories: the government can gather data first-hand, from the target itself, or second-hand, from a third party with whom the target shared its data. When the government gathers data first-hand, its activities likely constitute a search and the Fourth Amendment's protections against unreasonable searches and seizures apply.[166] When government gathers data second-hand from private parties, the Third-Party Doctrine says that the data-gathering is not a search and that the Fourth Amendment, for the most part, does not apply.[167]

First-hand data-gathering includes most traditional forms of police surveillance. Law enforcement regularly observes suspects with its own eyes and ears, can track a suspect's whereabouts by following him, and with the appropriate authorization, can search the suspect's home or property. Law enforcement can even surreptitiously record the activities of a suspect. And if all of those methods fail to provide enough information, the government can simply ask the suspect for information. Even then, in civil lawsuits between the Government and a private party, if the private party refuses to talk to the government, the government can force him,

---

[164] *Id.*

[165] Jason Shueh, *Big Data Could Bring Government Big Benefits*, GOV. TECH. (Mar. 21, 2014), http://www.govtech.com/data/Big-Data-Could-Bring-Governments-Big-Benefits.html.

[166] *See infra* Part IV.A.

[167] *See infra* Part IV.B. Technically, content data is still protected by the Fourth Amendment. This subtle yet important distinction is discussed more in Part IV.B.

her, or in the case of a business, it, to speak via use of a subpoena or, after a suit commences, depositions and even trial examination.[168]

Second-hand data-gathering requires the government to gather data from a private party for information about the activities of the suspect. The private parties can be human beings who witnessed the target say or do something. But private parties can also include machines or objects, owned by private parties that "witnessed" the target say or do something.[169] Google's, Sprint's, Facebook's, Apple's, and Twitter's information about the activities of people using their networks, phones, and computers is, to the government, second-hand data about a suspect's activities. Unlike information acquired by human witnesses, however, information acquired by computers and machines is not subject to the same kind of reliability concerns that plague human witnesses.[170] Unlike information stored in a human brain, data stored in a machine can be stored accurately for long periods of time and be easily combined with other data to paint a highly-detailed portrait of a suspect's activities.

---

[168] The Government may sue businesses or individuals civilly. Consumer protection is an example of such a lawsuit in which the State seeks data on an individual or business for the purposes of civil, not criminal prosecution. *See, e.g.*, OHIO REV. CODE ANN. § 1345.06 (West 2014) (giving the Ohio Attorney General the power to Subpoena information for civil consumer protection lawsuits). The Defendant of a civil suit can always assert their Fifth Amendment privilege against self-incrimination. But if the Defendant does, then the Defendant may lose the ability to defend themselves in the civil suit.

[169] *See* Smith v. Maryland, 442 U.S. 735, 736 n.1 (1979) (describing a pen register).

[170] *See* Laura Engelhardt, *The Problem With Eyewitness Testimony: Commentary on a Talk by Barbara Tversky & George Fisher*, 1 STANFORD JOURNAL LEGAL STUDIES 25 (1999), *available at* http://agora.stanford.edu/sjls/images/pdf/engelhardt.pdf (discussing problems of reliability with eyewitness testimony).

C. *The Size and Scope of the Government's Second-Hand Data-Collection Efforts*

While it is impossible to fully document the government's second-hand data-collection efforts, it is clear that the government regularly collects large amounts of data about its citizens from private parties, whether for criminal or civil investigations, or for national security purposes.

Facebook revealed that for a period of six months ending on December 31, 2012, federal, state, and local governments requested data on Facebook users between 9,000 and 10,000 times.[171] The Government inquired into the activities of 18,000 to 19,000 users.[172] Microsoft revealed over that same 6-month period, the government received between 6,000 and 7,000 national security warrants affecting between 31,000 and 32,000 users.[173] Google regularly reports government requests for information about its users.[174] It reports that government entities in the United States requested data about its users nearly 11,000 times.[175] This data does not include Foreign Intelligence Surveillance Court ("FISA court") orders, which Google says it cannot share.[176] Unlike Microsoft and Facebook, Google breaks down the means that the government uses to gather this data.[177] According to Google, 68% of U.S. government requests are subpoenas, 22%

---

[171] Press Release, Facebook, Facebook Releases Data, Including All National Security Requests (June 14, 2013) (on file with author), *available at* http://newsroom.fb.com/News/636/Facebook-Releases-Data-Including-All-National-Security-Requests.

[172] *Id.*

[173] Press Release, Microsoft, Microsoft's U.S. Law Enforcement and National Security Requests for Last Half of 2012 (June 14, 2013) (on file with author), *available at* http://blogs.technet.com/b/microsoft_on_the_issues/archive/2013/06/14/microsoft-s-u-s-law-enforcement-and-national-security-requests-for-last-half-of-2012.aspx.

[174] Richard Salgado, *Government Requests for User Information Double Over Three Years*, GOOGLE BLOG (Nov. 14, 2013), http://googleblog.blogspot.com/2013/11/government-requests-for-user.html.

[175] *Id.*

[176] *Id.*

[177] *Id.*

warrants, 6% other court orders, 2% pen registers, and 1% from emergency disclosure requests.[178]

Despite the reputational damage the NSA suffered due to the Edward Snowden leaks, the NSA is moving full-steam ahead with Big Data collection orders of magnitude larger than anything ever seen or done before. The NSA is currently constructing a 100,000 square foot, $1.2 billion data center in Utah. Maintenance costs alone are calculated at $20 million per year.[179] The center will employ 100 or so technicians and will allow NSA agents to connect remotely from anywhere in the world.[180] The data capacity is rumored to be 5 zettabytes, which is the data equivalent of 1.25 trillion DVDs.[181] Some estimates suggest the collection capacity reaches 1 yottabyte, or 250 trillion DVDs.[182] But some of these estimates are hard to believe because of just how staggeringly large that amount of data actually is, considering that global Internet traffic in the year 2015 is expected to reach only 1 zettabyte.[183]

But the government's thirst for second-hand Big Data, for all the reasons outlined by IBM and other Big Data companies, cannot be confined to investigations. States, for example, have been purchasing income data from credit bureaus like Equifax for some time now to help determine eligibility for state-run programs like welfare.[184]

---

[178] *Id.*

[179] Howard Berkes, *Booting Up: New NSA Data Farm Takes Root in Utah*, NPR (Sept. 23, 2013, 5:39 PM), http://www.npr.org/blogs/alltechconsidered/2013/09/23/225381596/booting-up-new-nsa-data-farm-takes-root-in-utah.

[180] *Id.*

[181] *Id.*

[182] *Id.*

[183] Thomas Barnett, Jr., *The Dawn of the Zettabyte Era*, CISCO BLOGS (June 23, 2011, 11:00 AM), http://blogs.cisco.com/news/the-dawn-of-the-zettabyte-era-infographic/.

[184] Hicken, *supra* note 157.

D. *The Many Ways Government Collects Second-Hand Data from Private Parties*

The government can obtain second-hand data from private parties in a variety of ways. First, the government can simply ask for it. According to Google, nearly 1% of requests for its user data from law enforcement are emergency requests.[185] A bill that has been proposed in Congress, called the Cyber Intelligence Sharing and Protection Act ("CISPA"), might dramatically increase this percentage. CISPA would make it legal for the government to ask companies for data about their customers and then protect those companies from lawsuits related to the handing over of that data, "notwithstanding any other provision of law."[186]

Second, the government can demand the data with a subpoena. A subpoena need not be reviewed or pre-approved by a court to be valid and enforceable.[187] Google says that 68% of its data requests from the government are in the form of a subpoena.[188] Subpoenas can request any information or documents that are at all relevant to an investigation. Relevance is defined very broadly and includes any information or documents that "might have the potential to lead to relevant information."[189] So long as a subpoena meets this very lenient standard, a court will deem the subpoena valid to the extent that the subpoena's demands are not overbroad or unduly burdensome.[190]

Third, the government can demand the information with a court order, which, by definition, does require prior approval by a

---

[185] Salgado, *supra* note 174.

[186] Declan McCullagh, *How CISPA Would Affect You (FAQ)*, CNET (Apr. 27, 2012, 4:00 AM), http://www.cnet.com/news/how-cispa-would-affect-you-faq/.

[187] *See, e.g.*, OHIO REV. CODE ANN. § 1345.06 (West 2014) (providing the attorney general with the authority to issue a subpoena without court approval).

[188] Salgado, *supra* note 174.

[189] *Responding to Subpoenas*, DIGITAL MEDIA LAW PROJECT, http://www.dmlp.org/legal-guide/responding-subpoenas (last visited June 1, 2014).

[190] *2nd Cir. Privilege Compendium: 4. Subpoena Not Overbroad or Unduly Burdensome*, REPORTERS COMMITTEE FOR FREEDOM OF THE PRESS http://www.rcfp.org/2nd-cir-privilege-compendium/4-subpoena-not-overbroad-or-unduly-burdensome (last visited June 1, 2014).

court.[191] Google says that 22% of its requests for data by the government are from warrants, and another 6% are from court orders.[192] The NSA collects much of its data by using secret FISA court orders, collecting huge sums of data from U.S. telephone companies, including AT&T, Verizon, and Sprint, and Internet service-providers like Facebook, Apple, Google, Microsoft, Yahoo, and AOL.[193] Statutes regulate these data-collection efforts.[194]

Fourth, the government can purchase the information. Big Data is valuable and companies are willing to sell.[195] For the right price,

---

[191] Hence the name, "court order."

[192] Salgado, *supra* note 174.

[193] *Everything You Need To Know About PRISM: A Cheat Sheet For the NSA's Unprecedented Surveillance Programs*, VERGE (July 17, 2013, 1:36 PM), http://www.theverge.com/2013/7/17/4517480/nsa-spying-prism-surveillance-cheat-sheet.

[194] Section 215 of the USA PATRIOT Act controls the government's ability to gather information from telecommunications companies. *See* USA PATRIOT Act of 2001, Pub. L. 107-56, § 215, 115 Stat. 272, 287 (2001) (codified as amended at 50 U.S.C. § 1861) ("Section 215"). Section 702 of the FISA Amendments Act regulates the government's ability to gather Internet data from Internet service-providers. *See* FISA Amendments Act of 2008, Pub. L. 113-103, § 702 (codified as amended at 50 U.S.C. § 1881a) ("Section 702"). Under Section 215, the government may obtain an order "requiring the production of any tangible things (including books, records, papers, documents, and other items) for an investigation to obtain foreign intelligence information not concerning a United States person or to protect against international terrorism or clandestine intelligence activities . . . ." *Id.* § 1861(a). When seeking authorization to obtain such an order from the FISA court, the government must also include "a statement of facts showing that there are reasonable grounds to believe that the tangible things sought are relevant to an authorized investigation." *Id.* § 1861(b)(2)(A). Section 702 of the FISA Amendments Act authorizes the government to, upon approval of the FISA Court, order an electronic communication provider to "immediately provide the Government with all information, facilities, or assistance necessary to accomplish the acquisition [of data]. . . ." *Id.* § 1881a(h)(1)(A). Acquisitions can be made without a court order if the government determines that exigent circumstances require immediate gathering of information important to national security. *Id.* § 1881a(c)(2).

[195] *See* Stephanie Faris, *Selling Big Data as a Service: 5 Industries Big Data Will Improve*, DATAVERSITY (Oct. 15, 2013), http://www.dataversity.net/selling-

government can access the same rich data-troves held by private organizations. For example, the federal government recently started buying access to a private database maintained by the credit bureau Equifax, called "The Work Numbers."[196] The database contains 54 million active salary and employment records and more than 175 million historical records from approximately 2,500 U.S. employers.[197] Equifax also sells this same data to credit card issuers, property managers, and auto lenders.[198]

Finally, the government can intercept the data using wiretaps, bugs, and Trojan horses among many other available tools. The NSA collects much of its data by tapping directly into telecommunications cables, both domestically and abroad.[199] These cables are owned by private-sector telecommunications companies, not the U.S. Government.[200] According to top-secret records provided by Edward Snowden, every day the NSA "Acquisitions Directorate" collects millions of records from Yahoo and Google this way.[201] Apparently, "[f]rom undisclosed interception points, the NSA . . . cop[ies] entire data flows across fiber-optic cables that carry information among the data centers of the Silicon Valley giants."[202] In just one month, the NSA had collected nearly 200 million new records, which included metadata and the content of text, audio, and video.[203] In a classic case of the pot calling the kettle black, a representative from Google blasted these activities,

---

big-data-as-a-service-5-industries-big-data-will-improve/ (discussing Big Data services as lucrative government contracts).

[196] Hicken, *supra* note 157.

[197] *Id.*

[198] *Id.*

[199] VERGE, *supra* note 193.

[200] *See Submarine Cable Map*, TELEGEOGRAPHY, http://www. submarinecablemap. com/#/submarine-cable/pacific-crossing-1-pc-1 (last visited June 1, 2014) (showing the many submarine cable lines and their private owners).

[201] Barton Gellman & Ashkan Soltani, *NSA Infiltrates Links To Yahoo, Google Data Centers Worldwide, Snowden Documents Say*, WASH. POST (Oct. 30, 2013), http://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html.

[202] *Id.*

[203] *Id.*

saying, "We are outraged at the lengths to which the government seems to have gone to intercept data from our private fiber networks . . . ."[204] A spokesperson for Yahoo remained more reserved, saying, "We have strict controls in place to protect the security of our data centers, and we have not given access to our data centers to the NSA or to any other government agency."[205] Google has since encrypted its dataflows between its data centers in an effort to secure its customers' data from the NSA's prying eyes.[206]

IV. THE LAW

Although private actors like Google are not limited by the requirements of the Fourth Amendment, the government is.[207] This distinction, however, does not mean much in the context of Big Data provided to the government by private parties. Under current Supreme Court doctrine, the government can acquire huge quantities of its citizens' data second-hand through private parties without triggering any of the protections of the Fourth Amendment.[208] There is only one significant limitation on this power: the Fourth Amendment's protections do apply to the *content* of communications.[209] But as Justice Sonia Sotomayor points out, this limitation does little to protect privacy when one's *non-content* data—metadata—already reveals so much.[210]

---

[204] *Id.*

[205] *Id.*

[206] *See* Chris Welch, *Google Encrypts Gmail Between Data Centers to Keep the NSA Out of Your Inbox*, VERGE (Mar. 20, 2014, 1:42 PM), http://www.theverge.com/2014/3/20/5530072/google-encrypts-gmail-between-data-centers-to-keep-out-nsa (discussing Google's encryption efforts).

[207] *See* United States v. Morrison, 529 U.S. 598, 621 (2000) (discussing the state action requirement).

[208] *See infra* Part IV.B.

[209] *See infra* Part IV.B.

[210] *See* United States v. Jones, 132 S. Ct. 945, 956 (2012) (Sotomayor, J., concurring).

A. *The Fourth Amendment*

The Fourth Amendment controls the legality of government searches and seizures. It reads:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.[211]

When it applies, the Fourth Amendment protects against unreasonable searches and seizures.[212] The guiding principle is simple—searches and seizures must be reasonable.[213] But because the Fourth Amendment does not define "unreasonable," the Supreme Court's Fourth Amendment search and seizure doctrine has become complicated and often times, counter-intuitive.

Under current Supreme Court doctrine, the government cannot search something without consent unless it has some degree of individualized suspicion that wrongdoing has occurred.[214] The degree of individualized suspicion required to search a particular suspect or his property increases as the suspect's expectation of privacy rises.[215] For example, a person having a loud conversation in the middle of a mall has no expectation of privacy. No degree of

---

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

[212] Terry v. Ohio, 392 U.S. 1, 9 (1968).

[213] *See id.* at 18 ("The distinctions of classical 'stop-and-frisk' theory thus serve to divert attention from the central inquiry under the Fourth Amendment—the reasonableness in all the circumstances of the particular governmental invasion of a citizen's personal security.").

[214] *See id.* at 30 (holding that a police officer must have more than a hunch but less than probable cause to be able to stop and question an individual suspected of wrongdoing). Furthermore, an officer conducting a so-called Terry Stop can only pat down the individual if the officer reasonably believes the individual possesses a weapon.

[215] *See, e.g.*, *id.* at 27 (permitting "frisks" when an officer stops a suspect on reasonable suspicion that the suspect has committed a crime, but for the sole purpose of ensuring officer safety); Kyllo v. United States, 533 U.S. 27, 31 (2001) (reiterating that entering the home generally requires a warrant); Bailey v. United States, 133 S. Ct. 1031, 1037 (2013) (discussing the "latitude" provided by the Fourth Amendment's reasonableness requirement in the context of searches incident to arrests);.

individualized suspicion is required before the government can listen to this person's public conversation. But if the suspect is at home, the suspect has a higher expectation of privacy. The degree of individualized suspicion required before the government can listen in on a conversation inside the home is, therefore, also higher. The question is, again, one of reasonableness. Would a person talking aloud in a mall "reasonably" expect privacy? No. Would a person talking to a family member at home? Yes.

Unlike the mall, searches of the home are presumptively unreasonable absent a warrant "particularly describing the place to be searched, and the persons or things to be seized" signed by a judge.[216] The home is typically the most protected realm,[217] often called one's "castle."[218] The only way for the government to search someone's home without a signed search warrant is for government to have probable cause that a suspect inside has committed a crime, plus some exigent circumstance that creates an immediate need for the government to enter the home (when there is not enough time to get a warrant).[219] Exigent circumstances might include the imminent destruction of evidence (e.g. flushing drugs down the toilet)[220] or the immediate impending physical harm to a person inside (e.g. domestic violence).[221]

The government is encouraged to take these requirements seriously when they apply because failure to abide by them will result in a court excluding evidence derived from the illegal search and seizure. Courts refer to this court-created policy of excluding

---

[216] Silverman v. United States, 365 U.S. 505, 511 (1961) ("At the very core stands the right of a man to retreat into his own home and there be free from unreasonable governmental intrusion.").

[217] *Id.* at 513 (Douglas, J., concurring) (discussing the warrant requirement's applicability to searches of the home).

[218] *See*, *e.g.*, *id.* at 511 n.4 (quoting William Pitt's description of the home as one's castle).

[219] Payton v. New York, 445 U.S. 573, 590 (1980) ("In terms that apply equally to seizures of property and to seizures of persons, the Fourth Amendment has drawn a firm line at the entrance to the house. Absent exigent circumstances, that threshold may not reasonably be crossed without a warrant.").

[220] United States v. Banks, 540 U.S. 31, 37–38 (2003).

[221] Brigham City, Utah v. Stuart, 547 U.S. 398, 403 (2006).

evidence obtained in violation of the Fourth Amendment as the "fruit of the poisonous tree" doctrine.[222] Evidence obtained illegally is the fruit born by the poisonous tree and, except in certain circumstances, courts will refuse to admit it at trial.[223]

Yet, the Fourth Amendment's protection against unreasonable searches and seizures does not apply if the Fourth Amendment itself does not apply. As any first-year law student knows, the Constitution does not apply to private actors, [224] with the exception of a few Amendments that are not relevant to privacy.[225] As a result, private businesses can "search and seize" all they want without violating the Constitution.[226] To be sure, searching and seizing a person's property without that person's consent can constitute theft.[227] It is not theft, however, when a private actor, such as a business searches and seizes data that a customer gives to it as part of the customer's relationship with that business.

The central question then is whether the Fourth Amendment applies to governmental acquisitions of data obtained second-hand through private parties. The answer depends on whether those governmental acquisitions constitute "searches" or "seizures." If the government conduct does not constitute a "search" or "seizure" within the meaning of the Fourth Amendment, then the Fourth Amendment does not apply to the government's conduct, even if reasonable Americans would think of that conduct as a search or seizure.[228]

---

[222] *See* Mapp v. Ohio, 367 U.S. 643, 657 (1961) (creating, explaining, and defending the Exclusionary Rule).

[223] *Id.*

[224] Burdeau v. McDowell, 256 U.S. 465, 475 (1921).

[225] *See* Steve Vladeck, *More Constitutional Curiosities: The Constitution and Private Conduct*, PRAWFSBLAWG (Nov. 21, 2005), http://prawfsblawg.blogs.com/prawfsblawg/2005/11/more_constituti.html (discussing the private-public distinction).

[226] *See* United States v. Francoeur, 547 F.2d 891 (5th Cir. 1977).

[227] *See, e.g.*, OHIO REV. CODE ANN. § 2913.02 (West 2014) (defining theft).

[228] *See* Kyllo v. United States, 533 U.S. 27, 31 (2001) (noting that whether the government's conduct constitutes a search is a question "antecedent" to applying the Fourth Amendment).

B. *Searches and Seizures*

The terms "search" and "seizure" are not defined in the Fourth Amendment or anywhere else in the Constitution.[229] Whether something is a search or seizure is not an intuitive exercise either.[230] What an ordinary citizen might call a search or seizure may or may not actually constitute a search or seizure under the Supreme Court's current Fourth Amendment jurisprudence. As a result, whether government conduct constitutes a search or seizure has been the subject of much litigation and has resulted in some blockbuster Supreme Court cases, most prominently, *Katz v. United States*[231] and *Jones v. United States*.[232]

In *Katz*, the Court held that the government's covert recording of the Defendant's phone-booth conversation constituted a search and seizure under the Fourth Amendment.[233] Though this holding is important, *Katz* is better known for its reasoning. According to the Court, the Fourth Amendment "protects people, not places."[234] In *Katz*, the government activity constituted a seizure because, in the Court's view, the Defendant had a justifiable expectation that his conversation would be private inside the phone booth, even though passersby could easily see inside.[235] Most importantly, the Court stated, "What a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection," but "what he seeks to preserve as private, even in an area accessible to the public, may be constitutionally protected."[236]

---

[229] To some, this indicates that these terms are meant to travel on the river of common law, changing as technology changes and society progresses. To others, the terms must be defined and confined to the meanings they had at the dawn of the Republic.

[230] *Kyllo*, 533 U.S. at 31 (saying this question is not "simple").

[231] 389 U.S. 347 (1967).

[232] 132 S. Ct. 945 (2012).

[233] *Katz*, 389 U.S. at 359.

[234] *Id.* at 351.

[235] *Id.* at 353.

[236] *Id.* at 351.

These words have come to be understood as creating a constitutionally protected "reasonable expectation of privacy,"[237] despite the fact that the phrase "reasonable expectation of privacy" cannot be found anywhere in the majority opinion.[238] The "reasonable expectation of privacy" standard actually comes from Justice John Marshall Harlan's concurrence, where he stated that his understanding of the Fourth Amendment was that "a person has a constitutionally protected reasonable expectation of privacy."[239] He further clarified that he understood "reasonable" to create a two-fold requirement: "first that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as 'reasonable.'"[240] Later Courts expressly adopted this subjective-objective, two-fold requirement.[241]

Most recently, *Jones* held that the clandestine attachment of a GPS tracking device to a defendant's vehicle constituted a search under the Fourth Amendment.[242] But the conclusion is not nearly as interesting as how the Court reached its conclusion. Although *Jones* was a 9-0 judgment, it can be read as having two different majority opinions based on two different legal theories.[243] One theory, advocated in Justice Antonin Scalia's majority opinion, is

---

[237] *Jones*, 132 S. Ct. at 952 (referencing the "*Katz* reasonable-expectation-of-privacy test").

[238] *Katz*, 389 U.S. at 349–59 (discussing privacy but never saying the words "reasonable expectation of privacy").

[239] *Id.* at 360 (saying, for the first time in the written decision, the words "reasonable expectation of privacy").

[240] *Id.* at 361.

[241] *See*, *e.g.*, California v. Ciraolo, 476 U.S. 207, 214 (1986) (holding that a defendant's expectation of privacy was "unreasonable" and was "not an expectation that society is prepared to honor"). *See also Jones*, 132 S. Ct. at 950 ("Our later cases have applied the analysis of Justice Harlan's concurrence in that case, which said that a violation occurs when government officers violate a person's 'reasonable expectation of privacy.'").

[242] *Jones*, 132 S. Ct. at 949.

[243] *See* Tom Goldstein, *Reactions to Jones v. United States: The Government Fared Much Better than Everyone Realizes*, SCOTUSBLOG (Jan. 23, 2012, 4:07 PM), http://www.scotusblog.com/?p=137698 ("I think that the correct way to understand the case is to read it as having two separate majority opinions.").

that the placement of the GPS tracker constituted a search because it involved a physical trespass onto the Defendant's property.[244] The other theory, advanced by Justice Alito's concurring opinion, argued that the GPS tracking was a search because it violated the Defendant's reasonable expectation of privacy.[245] But that five justices, including Justice Sotomayor, went further and applied *Katz* (or at least the Harlan concurrence) indicates that Justice Harlan's reasonable expectation of privacy standard is alive and well on the Court.[246]

Particularly relevant to government acquisitions of Big Data is Justice Sotomayor's concurrence. There, she expresses deep concern about the ability of long-term GPS monitoring to generate a "precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations."[247] In Justice Sotomayor's view, the efficient, low-cost nature of GPS data collection circumvents the ordinary constraints that, in practice, restrict the ability of law enforcement to monitor a person's whereabouts 24/7 over a long period of time.[248] Such constraints include limited resources and the generation of public antipathy toward such comprehensive monitoring practices.[249] Justice Sotomayor states that she would take into account the attributes of GPS technology when considering whether there is a reasonable

---

[244] *See Jones*, 132 S. Ct. at 949 (reasoning that the conduct constituted a search because law enforcement trespassed when it attached the GPS device to the Defendant's car).

[245] *Id.* at 964 (Alito, J., concurring); *see also id.* at 954 (Sotomayor, J., concurring) ("I join the Court's opinion because I agree that a search within the meaning of the Fourth Amendment occurs, at a minimum, '[w]here, as here, the Government obtains information by physically intruding on a constitutionally protected area.'").

[246] *See id.* at 954–55 (supporting use of Justice Harlan's reasonable expectation of privacy test to resolve case); *id.* at 958 (Alito, J., concurring) (same).

[247] *Id.* at 955 (Sotomayor, J., concurring).

[248] *See id.* at 956 ("And because GPS monitoring is cheap in comparison to conventional surveillance techniques and, by design, proceeds surreptitiously, it evades the ordinary checks that constrain abusive law enforcement practices . . . .").

[249] *Id.*

expectation of privacy "in the sum of one's public movements."[250] Specifically, she states that she "would ask whether people reasonably expect that their movements will be recorded and aggregated in a manner that enables the Government to ascertain, more or less at will, their political and religious beliefs, sexual habits, and so on."[251] She adds that the Fourth Amendment's goal is "to curb arbitrary exercises of police power to and prevent 'a too permeating police surveillance.'"[252]

C. *The Third-Party Doctrine*

The Harlan "reasonable expectation of privacy" test has given rise to a few additional doctrines defining when government activities would or would not violate a person's reasonable expectation of privacy and, therefore, constitute a search under the Fourth Amendment. One such doctrine is the so-called "Third-Party Doctrine," which the Supreme Court formally adopted in *United States v. Miller*.[253] There, the Court held that a law enforcement request to a bank for information about a bank customer's records did not constitute a search under the Fourth Amendment.[254] The Court held that a bank customer that knowingly reveals his affairs to another also knowingly takes the risk that the other will convey that information to the Government.[255] The Court then defined, for the first time, the Third-Party Doctrine:

> [T]he Fourth Amendment does not prohibit the obtaining of information revealed to a third party and conveyed by him to government authorities, even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed.[256]

In other words, the Fourth Amendment does not apply to information a person knowingly gives to another because

---

[250] *Id.*
[251] *Id.*
[252] *Id.*
[253] 425 U.S. 435 (1976).
[254] *Id.* at 439–40.
[255] *Id.* at 443.
[256] *Id.*

providing information to another destroys any reasonable expectation of privacy.[257] Put simply, the Fourth Amendment only protects those secrets that you keep to yourself.

Ever since *Miller*, the Third-Party Doctrine has been a cornerstone of Fourth Amendment jurisprudence upon which the Supreme Court has relied on multiple occasions to permit law enforcement to gather information without triggering the protections and limitations of the Fourth Amendment. Garbage bags placed on the street, even if opaque, are searchable without a warrant due to the Third-Party Doctrine.[258] Informants can wear a wire without triggering Fourth Amendment protections because of the Third-Party Doctrine.[259] Police officers can even fly helicopters 400 feet above someone's house, hover in one spot to achieve the perfect angle for viewing into the house, and then peer inside to record information, all without so much as reasonable suspicion, thanks to the Third-Party Doctrine.[260]

There is one important limitation on this "share information at your own risk"[261] Third-Party Doctrine. In *Smith v. Maryland*, the Court held that the warrantless recording of the telephone numbers that a man dialed from his home phone was not a search under the Third-Party Doctrine.[262] According to the *Smith* Court, a person has no reasonable expectation of privacy in the phone numbers she dials, even to a machine.[263] But at the same time, the Court also held that the individual does have a reasonable expectation of privacy in the *content* of the phone conversation itself.[264] The Court has upheld this distinction in other contexts as well. For example,

---

[257] Orin Kerr, *The Case for the Third-party Doctrine*, 107 MICH. L. REV. 561, 563 (2009).

[258] California v. Greenwood, 486 U.S. 35, 41 (1988).

[259] Lopez v. United States, 373 U.S. 427 (1963).

[260] Florida v. Riley, 488 U.S. 445 (1989).

[261] Kerr, *supra* note 257.

[262] Smith v. Maryland, 442 U.S. 735, 741–42 (1979).

[263] *See id.* at 745 ("We are not inclined to hold that a different constitutional result is required because the telephone company has decided to automate.").

[264] *See id.* at 741 (distinguishing between the contents of a telephone conversation, which is protected, and the non-content numbers the Defendant dialed, which it holds is not protected).

while a postal mailing's destination, sender, recipient, and any other contextual information readily viewable by the postal worker are not protected,[265] the content of the postal mailing is protected.[266] This content, non-content distinction therefore acts as a limitation on the ability of the Third-Party Doctrine to circumvent Fourth Amendment protections.

V. A BETTER APPROACH

The Third-Party Doctrine, though an invaluable tool in the twentieth century, must give way to a doctrine better tailored to the realities of American life in the twenty-first century. The Court should adopt the *Klayman v. Obama*[267] approach and hold that government acquisitions of data intrusive enough to cross the *Jones*[268] threshold are subject to protections of the Fourth Amendment.

A. *Time for a Change*

The world is not as it was in 1789 when the Bill of Rights was first written; or forty-six years ago when *Katz* was decided; or even thirty years ago, when the Court applied the Third-Party Doctrine to data collected by machines. In the Internet age, people's daily activities, some mundane, some incredibly private, occur online and almost exclusively through private third parties—mostly machines. The phone booths in *Katz* are largely a relic of the past, a vestige of the analog age long-since deceased. Americans now talk through cell phones and online chat services like Skype. Americans text and write emails instead of letters. Smartphones and their suite of apps, phone, GPS location, Internet, texting, and cloud storage have moved much of the core of human interaction online. Even human sexuality, in e-books, chat rooms, text

[265] United States v. Forrester, 512 F.3d 500, 511 (9th Cir. 2008) (citing Supreme Court and Ninth Circuit opinions to establish the Fourth Amendment differences between the content of a communication and its contextual information).

[266] *Ex parte* Jackson, 96 U.S. 727, 733 (1877).

[267] 957 F. Supp. 2d 1 (D.D.C. 2013).

[268] United States v. Jones, 132 S. Ct. 945 (2012).

messages, emails, and websites, has made its way online, in part because people believe that their online activities are private. Soon, nearly every device a person owns, from clothes and contact lenses to refrigerators, cars, and thermostats, will electronically monitor that person's daily life, inside the home and out, and even *inside the body and out*.[269]

And while the government could not directly gather this data without triggering Fourth Amendment protections,[270] under the Third-Party Doctrine the government can legally acquire the exact same data by merely asking private parties for it, or in some cases, intercepting it. Limitations on the Third-Party Doctrine, like the already flimsy content/non-content distinction offered by *Smith*, make very little difference in an age when one's metadata alone can be as revealing as the contents of a person's communications. Armed with mountains of highly-intrusive data obtained without Fourth Amendment scrutiny, the ability of government to perform sweeping, suspicionless searches on millions of Americans has never been easier, cheaper, more effective, and more worrisome.

At least one Justice would agree with this assessment. In *Jones*, Justice Sotomayor declared that in the information age, "it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in [the metadata] information voluntarily disclosed to third parties."[271] Justice Sotomayor explained that:

> This [Third-Party Doctrine] approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks. People disclose the phone numbers that they dial or text to their cellular providers; the URLs that they visit and the e-mail addresses with which they correspond to their Internet service providers; and the books, groceries, and medications they purchase to online retailers.[272]

She concluded that this issue need not be resolved at the time of the holding because a narrower ruling was sufficient to resolve

---

[269] *See supra* Part II.B.1.
[270] *See supra* Part IV.A.
[271] *Jones*, 132 S. Ct. at 957 (Sotomayor, J., concurring).
[272] *Id.*

the case.[273] But, like Justice Harlan's concurrence in *Katz*, she laid the groundwork for future doctrine.

Whether four more Justices would agree with Justice Sotomayor's proposal is unknown. Justice Alito's concurrence at least suggests that he and at least three other Justices are amenable to limiting the application of the Third-Party Doctrine when it comes to highly-intrusive, comprehensive data-collection techniques like the one in *Jones*.[274] After all, if the long-term collection of GPS data—data about a person's public whereabouts—triggers the Fourth Amendment protections despite the seeming applicability of the Third-Party Doctrine, then so too should government acquisitions of rich troves of Big Data. After all, Big Data likely contains comprehensive public and private GPS data *in addition to* copious amounts of personal data.

More fundamentally, the Third-Party Doctrine is premised on the flawed assumption that knowing exposure to others necessarily defeats any reasonable expectation of privacy. Although a court may declare that someone's public whereabouts, in-store behavior and purchases, smartphone activity, and website browsing activity are "knowingly exposed to others," it is not necessarily true that the individual did so without a reasonable expectation of privacy. Sharing information with one person does not equate to sharing information with the world. This is especially true when the "person" receiving the data is a hard drive belonging to a large corporation like Time Warner Cable, Google, or Microsoft—companies that boast about their robust "privacy" policies and promise to protect people's data.[275]

Sometimes, a person's data shared with another cannot be reasonably defined as having been "knowingly exposed." British

---

[273] *Id.*

[274] *See id.* at 964 (Alito, J., concurring) (holding that people have a reasonable expectation of privacy in highly intrusive data, like comprehensive GPS records).

[275] *See, e.g.*, *Your privacy is our priority*, MICROSOFT, http://www.microsoft.com/security/online-privacy/overview.aspx (last visited Aug. 9, 2014) (stating that "your privacy" is Microsoft's "priority" and expressing its "longstanding commitment" to privacy).

"tech-blogger" Doctorbeet did not know that he was sharing his television-watching activity with a third party.[276] The woman who sends her boyfriend a text from an abortion clinic does not "knowingly expose" her location to the rest of the world.[277] Neither does the user who searches "Oklahoma abortion law" or "Hillary Clinton naked gay overbite porn."[278] Nor do most people who send volumes of hidden metadata along with the content of their communications expect that data to be made available to the public.

Sometimes people "knowingly expose their data" involuntarily. Doctorbeet never intended, before or after he discovered that his TV was snooping on him, to share his television-watching activity with others. In fact, he continued to "knowingly expose" his data after commanding his television to stop snooping. Arguably, the private collection of people's data is almost always involuntary. No one, or at least very few, would volunteer their data to others. If people had a choice to turn off data collection, odds are that they likely would.[279] This is especially true when the data reveals intimate and secret details of a person's life. Indeed, a recent poll showed that 86% of Internet users have "taken steps online to remove or mask their digital footprints."[280]

Furthermore, the argument that people would freely give up their privacy in exchange for quality online services is belied by facts and life experience. Very few companies provide services free of tracking, even when consumers pay for those services.[281] For the vast majority of services, consumers have no other choice but to hold their nose, click "I accept," and subject themselves to

---

[276] *See supra* Part II.B.3.

[277] *See supra* Part II.E.

[278] *See supra* Part II.E.

[279] *See* Rainie et al., *supra* note 141 (reporting the results of a poll indicating the pro-privacy attitudes of the majority of Internet users).

[280] *See id.*

[281] *See, e.g.*, *Privacy Policy*, PANDORA, http://www.pandora.com/privacy (last visited Oct. 6, 2014) (permitting advertisers to collect cookie and beacon-based data from Pandora listeners, regardless of whether the listener is a paying customer).

even more data collection. Saying that Americans have a "choice," when the only viable alternative is to opt out of all twenty-first century technology, is to say that American's have a Hobson's choice, or really, no real choice at all. Americans should not have to shed the protections of the Fourth Amendment to enjoy the benefits of the information age.

All of these situations undermine the rote application of the Third-Party Doctrine to people's data. Proving that information was knowingly exposed to a third party, although important to the government's argument, is, by itself, insufficient to establish that the data that the government seeks to acquire is unprotected by the Fourth Amendment. Indeed, all of the data in *Jones* was "knowingly exposed to others." In fact, that data was not only knowingly exposed to others, it was knowingly exposed directly to the public when the Defendant drove around on public streets. This is an important distinction between the GPS data in *Jones* and the kind of Big Data at issue here. Yet, even in *Jones*, the individual retained a reasonable expectation of privacy in the totality of his movements despite exposing each of those movements directly to the public. Certainly, if it can be reasonable to expect privacy in information that a person knowingly exposes directly to the public, then it can be reasonable to expect privacy in information that a person knowingly exposes to a private third party. And if it can be reasonable to expect privacy in public movements, then there is no doubt that someone can expect privacy in information that he or she unwittingly or involuntarily exposes to the public.

B. *The* Klayman v. Obama *Approach*

In *Klayman v. Obama*,[282] a federal district court granted a preliminary injunction against the NSA's bulk collection and analysis of telephone call metadata after concluding that *Smith*—the case giving rise to the Third-Party Doctrine—was so old and unlike the facts before the court that it was simply inapposite.[283]

---

[282] 957 F. Supp. 2d 1 (D.D.C. 2013).
[283] *Id.* at 43.

The court concluded that, "In sum, the *Smith* pen register[284] and the ongoing NSA Bulk Telephony Metadata Program have so many significant distinctions between them that I cannot possibly navigate these uncharted Fourth Amendment waters using as my North Star a case that predates the rise of cell phones."[285]

Moreover, the *Klayman* court did not just limit the application of the Third-Party Doctrine. It also supplied an alternative: the *Katz* "reasonable expectation of privacy" test.[286] Though the court declined to rule on the constitutionality of the NSA's warrantless surveillance program, it did grant a temporary injunction against the program on the theory that the program was substantially likely to violate people's reasonable expectation of privacy.[287] By employing this reasoning, the court fully embraced the Alito-concurrence opinion in *Jones*: Highly intrusive data collection by government can violate the Fourth Amendment even if each individual data point in that data set was knowingly exposed to others.[288]

By concluding that *Smith* is too inapposite to bind its judgment, applying the *Katz* "reasonable expectation of privacy" test instead, and explicitly embracing Alito's concurrence in *Jones*, the *Klayman* court shows the Supreme Court how Fourth Amendment jurisprudence can be adapted to the twenty-first century in a way that does not damage precedent or the Fourth Amendment itself.

C. *Courts Should Adopt the* Klayman *Approach*

This Article suggests courts adopt the *Klayman* approach.[289] Courts should decline to mechanically apply the Third-Party Doctrine to data the government obtains from third parties, and instead hold that Americans have a reasonable expectation of

---

[284] A pen register is a device employed by a telephone company that records the numbers dialed on a telephone. *See* Smith v. Maryland, 442 U.S. 735, 736 n.1 (1979) (describing a pen register).

[285] *Klayman*, 957 F. Supp. 2d at 37.

[286] *Id*. at 32–37.

[287] *Id.* at 37.

[288] *See id.* at 30–36 (abiding by the *Jones* approach and frequently citing *Jones*).

[289] *See generally id.*

privacy in the collection of any amount of data, including non-content metadata, that is as comprehensive and intrusive as the GPS data collected in *Jones*. This would mean that any data collection that is intrusive enough to cross the *Jones* threshold would constitute a search under the Fourth Amendment and subject that data-collection to the reasonableness requirements of the Fourth Amendment, regardless of whether the Third-Party Doctrine applies. This would most certainly include, and therefore protect against, government acquisitions of Big Data.

This fix would not undermine law enforcement because it would still permit the government to obtain incriminating data on persons suspected of wrongdoing. It would leave intact the Third-Party Doctrine while at the same time dramatically limiting its application. In other words, the Third-Party Doctrine would still apply, but only up to a point. Within the narrow window between zero data and the *Jones* intrusiveness threshold, the *Klayman* and *Jones* approach would enable government to collect data without Fourth Amendment scrutiny. Above the *Jones* threshold, where the data is so revealing and so intrusive as to create a reasonable expectation of privacy in the data, collecting the data would constitute a search or seizure under the Fourth Amendment and be subject to Fourth Amendment protections, even if it would not have been under the Third-Party Doctrine as defined in *United States v. Miller*.[290] And yet, even when the acquisition of the data would cross the *Jones* threshold and constitute a search or seizure under the Fourth Amendment, the Government could still get that essential data if it could first articulate at least some degree of suspicion that the individual committed or is committing some crime.[291] This is standard procedure for law enforcement in any other context, and it has been for a very long time.[292] Applying *Jones* to Big Data simply extends the venerable protections of the Fourth Amendment to our information age.

---

[290] *See* United States v. Miller, 425 U.S. 435 (1976).

[291] Terry v. Ohio, 392 U.S. 1, 30 (1968) (upholding "reasonable suspicion" as the suspicion level required to stop a suspect).

[292] *See, e.g.*, *id.* (upholding reasonable suspicion standard in 1976).

When the Exclusionary Rule, colorfully called the "fruit of the poisonous tree doctrine," became law, the government was forced to self-regulate. Similarly, a holding that *Jones* applies to Big Data collection as well as to GPS data-collection would force the government to self-regulate. The more data the government collects, the harder the government will apply the brakes to its own data collection to ensure that it does not cross the *Jones* threshold. This is because the line between acceptable data-collection and the kind of highly-intrusive data-collection at issue in *Jones* is not entirely clear, at least not yet. As long as the government is interested in prosecuting the target of its investigation and data-collection efforts, government will self-regulate out of a fear that it might go too far, cross the *Jones* threshold, and trigger Fourth Amendment protections.[293] The cost to law enforcement of unintentionally going too far without abiding by the limitations of the Fourth Amendment is high; the prosecution cannot, barring a few exceptions, introduce such evidence at trial.[294]

Just how much data collection it would take to cross the *Jones* threshold is, therefore, the big question. Databases of Big Data can

---

[293] The *Mapp* Exclusionary Rule is very powerful. *See* Mapp v. Ohio, 367 U.S. 643, 657 (1961) (creating the Exclusionary Rule). So long as law enforcement's goal is to prosecute and convict a suspect, then law enforcement will take care to ensure that its searches and seizures are legal and any evidence obtained pursuant to those searches and seizures will survive a motion to suppress. *See* Bradley C. Canon, *Is the Exclusionary Rule in Failing Health? Some New Data and a Plea Against a Precipitous Conclusion*, 62 KY. L.J. 681 (1973–74) (documenting the success of the Exclusionary Rule in deterring illegal police conduct). *But see* Thomas Y. Davies, *Critique, On the Limitations of Empirical Evaluations of the Exclusionary Rule: A Critique of the Spiotto Research and* United States v. Calandra, 69 NW. U. L. REV. 740 (1974) (arguing that it was impossible to determine whether the Exclusionary Rule had a deterrent effect); United States v. Janis, 428 U.S. 433, 446 (1976) (stating that the Court's debate on the Exclusionary Rule "has been unaided, unhappily, by any convincing empirical evidence on the effects of the rule").

[294] United States v. Leon, 468 U.S. 897, 910 (1984). This assumes, as does the rationale behind the Exclusionary Rule itself, that law enforcement cares about prosecutions. True, if law enforcement does not care about the prosecutions and cares about something else, like intimidation, then this rule would have no effect.

paint a highly-detailed portrait of a person. It takes very little data to reveal a great deal of sensitive, private, and even intimate information about an individual. Even more than GPS data, Big Data can reveal a great deal about a person's sexual preferences, health issues, legal questions, religious beliefs, etc.[295] Conceivably, even a very limited, granular dataset of pure metadata could cross the *Jones* threshold (think of the example of the woman buying batteries at a sex shop or the woman sending a text message from an abortion clinic). This at least means that the government would not be able to freely amass huge databases of Big Data on thousands or millions of people without triggering the protections of the Fourth Amendment. The *Klayman* court has already held that the NSA's bulk collection of five-year's-worth of telephony metadata from hundreds of millions of people without a warrant likely violates *Katz* and *Jones*.[296]

Applying the Fourth Amendment to government acquisitions of Big Data would prevent government from acquiring massive databases on thousands of people for another reason: The Fourth Amendment requires individualized, not generalized, suspicion.[297] The appropriate level of suspicion that Person A committed a crime would grant law enforcement the right to gather data about Person A, not Person B, or anyone else for that matter. Government would have to suspect that each and every person in the dataset that it seeks to acquire committed a crime, or else it would give up the right to use any incriminating data it did find on an unsuspected person in that database. That a database "happens" to contain other information on other people is no excuse either. In a modern database, only basic database manipulation skills are required to separate one individual's data from others. Surely, the government entities or private parties that maintain databases

[295] *See supra* Part II.D.

[296] Klayman v. Obama, 957 F. Supp. 2d 1, 30–32 (D.D.C. 2013). *But see* ACLU v. Clapper, 959 F. Supp. 2d 724 (S.D.N.Y. 2013) (upholding as constitutional the same NSA program).

[297] Terry v. Ohio, 392 U.S. 1, 30 (1968) (holding that a police officer must have more than a hunch but less than probable cause to be able to stop and question an *individual* suspected of wrongdoing).

would have the minimal level of technical expertise necessary to exclude unsuspected persons and their data from government inquiries into suspected persons.

Such a holding would comport with precedent, abide by the reasonableness standard of the Fourth Amendment, and constrain the Court to its more limited role as interpreters of the law rather than creators of it. Justice Sotomayor's concurrence in *Jones* all but pre-writes this future opinion. Justice Sotomayor need only change "GPS data" to "Big Data," because the reasoning of the two opinions would be identical. She concurred in *Jones* because GPS monitoring generates a "precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations."[298] At the very least, the exact same can be said of government acquisitions of Big Data. More realistically, Big Data is even more revealing and potentially more damaging than mere GPS data of a person's public whereabouts. As shown by the AOL's disastrous release of its customers' "anonymized" search data, even a small quantity of data—even metadata—can reveal a great deal about a person, especially because a great many, if not most, people believe that their Internet activities are private or anonymous.[299] The prevalence of Internet pornography, and even arguably the success of erotic e-books like *Fifty Shades of Grey*,[300] demonstrate that many, if not most, people believe, even if erroneously, that their electronic activities are private, or at least anonymous. This is only reinforced by the two-way mirror that prevents people from seeing the myriad faceless entities that are watching people's every move. Surely, if the totality of a person's public whereabouts creates a reasonable expectation of privacy, then so too does a data trove containing the most intimate details of a person's life.

Perhaps most importantly, subjecting government acquisitions of Big Data to Fourth Amendment scrutiny does not undermine the

---

[298] United States v. Jones, 132 S. Ct. 945, 955 (2012) (Sotomayor, J., concurring).
[299] *See* Rainie et al., *supra* note 141.
[300] E. L. JAMES, FIFTY SHADES OF GREY (Vintage 2012).

government's ability to protect its citizens from terrorist attack or prevent the government from using Big Data to improve government services. As effective as the Exclusionary Rule may be at encouraging the government to comply with the Fourth Amendment in most situations, the Exclusionary Rule has no effect when the purpose of the government's search or seizure is something other than prosecution.[301] Prosecution is not the government's goal when it comes to thwarting an imminent terrorist attack. Nor is it the government's goal when it comes to improving government services. In these situations, the government can continue to gather all the Big Data it desires without worrying about the consequences of violating a citizen's Fourth Amendment rights. This may seem like a gaping loophole in the protections of the Fourth Amendment, but in these limited situations, such a loophole serves as a venting mechanism to ensure that protecting American citizens' Fourth Amendment rights does not interfere with the government's ability to keep its citizens safe or improve government services: security, liberty, privacy, and governmental efficacy, all at the same time.

To be sure, this Article's recommendation is not perfect. Just like *Jones*, it would create serious line-drawing problems—the kind of "thorny" issues that the Scalia majority wanted to avoid in *Jones*.[302] It is unclear precisely when the government's data collection becomes too intrusive, crosses the *Jones* threshold, and becomes subject to the protections of the Fourth Amendment. At some point, the Court will also need to determine the appropriate level of suspicion necessary to obtain intrusive data without a warrant and when, if ever, the government must first obtain a warrant.

These important questions must be answered at some point. But just like *Jones*, the courts need not answer all of these questions immediately.[303] As with any judicially created doctrine, courts have the ability to develop the jurisprudence over time and

---

[301] *See supra* note 294.
[302] *Jones*, 132 S. Ct. at 954 (majority opinion).
[303] *Id.* at 964 (Alito, J., concurring).

iron out the wrinkles as they arise. For now, courts can take the first step of requiring government to abide by the rigors and protections of the Fourth Amendment whenever government seeks to acquire Big Data, either on its own or from private parties. Otherwise, the status quo—a not-so-subtle circumvention around the Fourth Amendment that effectively outsources government surveillance to private parties—will continue to threaten the freedom and liberty of Americans, chill speech, and shift the expectations of Americans further toward Orwell's *1984*.

VI. CONCLUSION

In an age when people's lives are constantly tracked, recorded, analyzed, and shared by private parties, the doctrine holding that "information knowingly exposed to private parties is unprotected by the Fourth Amendment," now threatens to swallow whole the privacy guaranteed by the Fourth Amendment. This Article suggests courts adopt the *Klayman v. Obama* approach and hold that the Fourth Amendment's protections apply to government acquisitions of Big Data. More specifically, courts should follow Justice Alito's reasoning in *United States v. Jones* to hold that government acquisitions of Big Data are searches subject to the reasonableness requirements of the Fourth Amendment. Surely, if the government's collection of simple GPS data in *Jones* was intrusive enough to constitute a search, then so are government acquisitions of Big Data. Though such a holding would leave unresolved many important questions, such as whether the collection of bulk data would require a warrant, it would be an important first step that would bring the Fourth Amendment into the twenty-first century and enable the next generation of Americans to conduct their lives without fear of unreasonable government searches and seizures of their most private data.