WELCOME TO THE MACHINE: PRIVACY AND WORKPLACE IMPLICATIONS OF PREDICTIVE ANALYTICS

Robert Sprague*

Cite as: Robert Sprague, *Welcome to the Machine: Privacy and Workplace Implications of Predictive Analytics*, 21 RICH. J.L. & TECH. 12 (2015), http://jolt.richmond.edu/v21i4/article12.pdf.

> Privacy is not simply an absence of information about us in the minds of others; rather it is the *control* we have over information about ourselves.[1]

> [T]he volume of information that people create themselves—the full range of communications from voice calls, e-mails and texts to uploaded pictures, video, and music—pales in comparison to the amount of digital information created *about* them each day.[2]

* J.D., M.B.A. Associate Professor of Legal Studies in Business, University of Wyoming College of Business, Department of Management & Marketing. A working draft of this article was presented at the Legal and Ethical Dimensions in Predictive Data Analytics Colloquium sponsored by Virginia Tech University and the Center for Business Information Analytics at the Pamplin College of Business, June 20, 2014. The author thanks Kellsie Jo Nienhuser, J.D. 2015 University of Wyoming College of Law, for her excellent research assistance.

[1] Charles Fried, *Privacy*, 77 YALE L.J. 475, 482 (1968).

[2] EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 2 (2014) [hereinafter BIG DATA: SEIZING OPPORTUNITIES], *available at* http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf, *archived at* http://perma.cc/9PNW-JBEK (emphasis in original).

I. INTRODUCTION

[1] Predictive analytics use a method known as data mining to identify trends, patterns, or relationships among data, which can then be used to develop a predictive model;[3] in many cases attempting to predict behavior. The advent of ubiquitous monitoring and tracking—from self-generated content, web browsing, online transactions, geolocation tracking, and infrastructure sensors—provide the "big data" needed for data mining and predictive analytics. Privacy law has not kept up, particularly since most of the data are "public", in that they are not secret or confidential. Yet, big data mining can reveal intimate facts and portrayals of individuals.

[2] After providing general background on data analytics, this article explores possible theories of privacy protection for predictive analytics; specifically under the evolving "mosaic" theory that has so far been considered, to varying degrees, in Fourth Amendment search scenarios. This article makes an argument that predictive analytics are ripe for privacy protection based on the mosaic theory.

[3] This article then turns to predictive analytics in the workplace, exploring ways in which big data is used in the employment context and considering what level of privacy protection may be available to workers. The conclusion is not optimistic, from a privacy advocacy perspective, as workplace privacy was already minimized before big data even made its appearance. For all practical purposes, it is impossible to avoid "emitting" digital information that can be collected, stored, analyzed, and used for a myriad of decision scenarios; all one can really do is be aware that it is

[3] *See* CHARLES NYCE, PREDICTIVE ANALYTICS WHITE PAPER 1 (2007), *available at* https://web.archive.org/web/20140305101843/http://www.theinstitutes.org/doc/predictivemodelingwhitepaper.pdf, *archived at* https://perma.cc/XR7P-DYPP; *see also* HERMAN T. TAVANI, ETHICS AND TECHNOLOGY: CONTROVERSIES, QUESTIONS, AND STRATEGIES FOR ETHICAL COMPUTING 150 (4th ed. 2013) (describing data mining as involving "the indirect gathering of personal information through an analysis of implicit patterns discoverable in data[;]" noting further that "[d]ata mining activities can generate new and sometimes non-obvious classifications or categories").

occurring—just about everywhere, just about all the time.

II. PREDICTIVE ANALYTICS

[4] Predictive analytics enable organizations to determine trends and relationships that may not have otherwise been readily apparent.[4] Increasingly sophisticated statistical models coupled with the growth of "big data"[5] have led to an increasing use of predictive analytics in a variety of situations.[6] The range of predictive analytics is bolstered by the

[4] *See* NYCE, *supra* note 3, at 1; *see also* CTR. FOR INFO. POLICY LEADERSHIP, HUNTON & WILLIAMS LLP, BIG DATA AND ANALYTICS: SEEKING FOUNDATIONS FOR EFFECTIVE PRIVACY GUIDANCE 1 (2013) [hereinafter BIG DATA AND ANALYTICS], *available at* http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf, *archived at* http://perma.cc/NAN3-79N6 ("While traditionally analytics has been used to find answers to predetermined questions, its application to big data enables exploration of information to see what knowledge may be derived from it, and to identify connections and relationships that are unexpected or were previously unknowable.").

[5] *See* BIG DATA AND ANALYTICS, *supra* note 4, at 1 (describing big data as "vast stores of information gathered from traditional sources (e.g., public record data, health data, financial and transactional data) and from new collection points . . . (e.g., web data, sensor data, text data, time and location data and data gleaned from social networks)"). "Big data is characterized by the variety of its sources, the speed at which it is collected and stored, and its sheer volume." *Id. See also* Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 74 (2013) (referring to big data as "novel ways in which organizations, including government and businesses, combine diverse digital datasets and then use statistics and other data mining techniques to extract from them both hidden information and surprising correlations"). Viktor Mayer-Schönberger and Kenneth Cukier, while noting there is no "rigorous" definition of big data, refer to it as "things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more." VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 6 (2013).

[6] *See, e.g.*, NYCE, *supra* note 3, at 1 (discussing the increasing use of analytics in insurance underwriting); *see also* BIG DATA AND ANALYTICS, *supra* note 4, at 1 (noting that analytics can "help identify individuals in need of social services, detect fraud, predict the effects of natural disasters, recognize patterns in scientific research and discover trends in consumer demand"); Steve Lohr, *Banking Start-Ups Adopt New Tools*

vast amount of increasingly available data: online transaction records, e-mail messages and metadata,[7] images, web browsing logs, search queries, health records, social networking interactions, geolocation tracking, and sensors deployed in infrastructure such as communications networks, electric grids, global positioning satellites, roads and bridges, as well as in homes, clothing, and mobile phones.[8] One can think of predictive analytics another way: "instead of people using search engines to better understand information, search engines will use big data to better understand people."[9]

---

for Lending, N.Y. TIMES, Jan. 19, 2015, http://www.nytimes.com/2015/01/19/technology/banking-start-ups-adopt-new-tools-for-lending.html, *archived at* http://perma.cc/L4JK-6HWE (noting that with the use of predictive analytics, loan underwriting could depend on whether customers use only capital letters when filling out forms).

[7] Metadata are essentially data about data. *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE xi (2014) [hereinafter BIG DATA AND PRIVACY], *available at* http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf, *archived at* http://perma.cc/L3CU-RCD8 ("Metadata are ancillary data that describe properties of the data such as the time the data were created, the device on which they were created, or the destination of a message.").

[8] *See, e.g.*, MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 83–97; Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 240 (2013); *see also* Ariana Eunjung Cha, *'Smart Pills' with Chips, Cameras and Robotic Parts Raise Legal, Ethical Questions*, WASH. POST (May 24, 2014), http://www.washingtonpost.com/national/health-science/smart-pills-with-chips-cameras-and-robotic-parts-raise-legal-ethical-questions/2014/05/24/6f6d715e-dabb-11e3-b745-87d39690c5c0_story.html, *archived at* http://perma.cc/WD2Y-6VS9 (describing how "companies and academic research teams are rushing to make ingestible or implantable chips that will help patients track the condition of their bodies in real time and in a level of detail that they have never seen before").

[9] Ian Kerr & Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, 66 STAN. L. REV. ONLINE 65, 65–66 (2013), http://www.stanfordlawreview.org/sites/default/files/online/topics/66_StanLRevOnline_65_KerrEarle.pdf, *archived at* http://perma.cc/CE3G-SM2Q.

[5] But predictive analytics can go a step further than traditional data analysis—creating a picture of social behavior that was not previously possible.[10] Ericka Menchen-Trevino notes that a new interdisciplinary field, computational social science, is forming around the social analysis of digital imprints left by e-mail, text messages, tweets, surfing the web, social media applications, and smart phones.[11] These data are not necessarily tracking transactional records of atomized behavior—such as the purchasing history of customers—but are keeping track of communication dynamics and social interactions.[12] For computational social scientists, big data is "big" not because of its size, but because its analytical potential is qualitatively different.[13] Indeed, some researchers claim that big data can track human behavior more precisely than theoretical models.[14] Big data can help illuminate the complexity that

[10] *See* Ericka Menchen-Trevino, *Collecting Vertical Trace Data: Big Possibilities and Big Challenges for Multi-Method Research*, 5 POL'Y & INTERNET 328, 328 (2013); *see also* BEN WABER, PEOPLE ANALYTICS 10–12 (2013) (discussing a "Sociometer" that incorporates "the critical sensors necessary to understand many aspects of human behavior"); Emily Singer, *Is "Self-Tracking" the Secret to Living Better?*, MIT TECH. REV. (June 9, 2011), http://www.technologyreview.com/view/424252/is-self-tracking-the-secret-to-living-better/, *archived at* http://perma.cc/3NMD-AXLK (reporting on the self-tracking movement which utilizes wireless sensing devices and smartphones to track personal lifestyle data).

[11] *See* Menchen-Trevino, *supra* note 10, at 328; *see also* Sandra González-Bailón, *Social Science in the Era of Big Data*, 5 POL'Y & INTERNET 147, 148 (2013) ("[W]hat makes Big Data unique is their higher level of detail and refinement in the quality of observations, not just the number of data points or the amount of memory that their storage takes.").

[12] *See* González-Bailón, *supra* note 11, at 148.

[13] *See* Menchen-Trevino, *supra* note 10, at 329.

[14] *See, e.g.*, González-Bailón, *supra* note 11, at 147–48 (discussing "end of theory" arguments); *see also* MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 70 ("In the future, our understanding will be driven more by the abundance of data rather than by hypotheses."); Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 23, 2008), http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory, *archived at*

interactions add to social dynamics "with an impressive level of detail."[15]

[6] Viktor Mayer-Schönberger and Kenneth Cukier provide a brief analysis of the promise—and the peril—of big data predictive analytics. Prior to big data, analytics relied on determining whether an individual was part of a group; for example, actuarial tables indicate that men over fifty years of age are more prone to colon cancer, so all men over fifty may pay more for health insurance.[16] In contrast, big data analysis is non-causal, identifying individuals—rather than groups—from a vast array of data.[17] Mayer-Schönberger and Cukier argue that, on the plus side, this makes profiling much more accurate, less discriminatory, and more individualized.[18] For example, rather than identifying an individual as a terrorist threat due to his or her nationality or religion, additional data points such as body language and other physiological patterns can be analyzed to make a more accurate determination of a possible threat.[19] On the down side, it may lead some to predict behavior based on mere

http://perma.cc/Z4UD-6X4L (claiming that "faced with massive data, this approach to science—hypothesize, model, test—is becoming obsolete"). "We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot." *Id.*

[15] González-Bailón, *supra* note 11, at 148.

[16] *See, e.g.*, MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 160.

[17] *See id.*

[18] *See id.* at 161.

[19] *See id.* at 159–61*; see generally* Matthew L. Jensen et al., *Automatic, Multimodal Evaluation of Human Interaction*, 19 GROUP DECISION & NEGOTIATION 367 (2010) (outlining an approach for automatically extracting behavioral indicators from video, audio, and text and exploring the possibility of using those indicators to predict certain human-interpretable judgments); Thomas O. Meservy et al., *Deception Detection Through Automatic, Unobtrusive Analysis of Nonverbal Behavior*, 20 IEEE INTELLIGENT SYS. 36, 36 (2005) (discussing the theory underlying an automated system that can infer deception or truthfulness from a set of features extracted from head and hands movements in a video).

probabilities; big data analytics can only "predict that for a specific individual, a particular future behavior has a certain probability."[20]

[7] Predictive analytics are not perfect. While they may reveal hidden correlations, there may be no causation. For example, Google engineers found a correlation between Google flu-related searches and outbreaks of the flu, identifying flu outbreaks before the Centers for Disease Control.[21] However, the engineers did not examine what caused those searches. For example, a few years later, Google's predictive capabilities came into question when it drastically overestimated peak flu levels based on search queries, most likely because Google's algorithms did not sufficiently take into consideration people who were not suffering from the flu conducting flu-related searches due to higher than usual press coverage of a flu outbreak.[22] "Imputing true causality in big data is a research field in its infancy."[23] In addition, while there may be a lot of data, they are not always complete or accurate and may contain outliers—all of which can lower the performance of data mining algorithms.[24]

III. PREDICTIVE ANALYTICS AND PRIVACY

[8] In the early years of America as a colony and a nation, privacy was

[20] MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 161. Mayer-Schönberger and Cukier remind us that numbers are far more fallible than we sometimes think. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 163.

[21] *See* Jeremy Ginsberg et al., Letter to the Editor, *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012, 1012–13 (2009) (arguing it is possible to use search queries to detect influenza epidemics in areas with a large population of web search users).

[22] *See* Declan Butler, *When Google Got Flu Wrong*, 494 NATURE 155, 155 (2013) (arguing that mining web and social media data for flu-tracking can only compliment, not substitute for, traditional epidemiological surveillance networks).

[23] BIG DATA AND PRIVACY, *supra* note 7, at 25.

[24] *See id.*

a relatively minor social concern in light of social norms. Church elders would regularly visit the homes of parishioners to ensure proper living;[25] family members, as well as visiting guests, would often sleep in the same beds;[26] and Henry Ford would send his "sociological investigators" to the homes of workers to ensure proper living before extending a wage bonus.[27] America's open frontier provided its own natural solitude.[28] Privacy was primarily limited to admonishing eavesdroppers—those who would stand outside the open eaves of a home and listen to the conversations within.[29]

[9] When Samuel Warren and Louis Brandeis proposed a "right to be let alone" in their seminal article *The Right to Privacy*,[30] it was in reaction to new intrusive technologies: "[i]nstantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that 'what is whispered in the closet shall be proclaimed

[25] *See* ROBERT ELLIS SMITH, BEN FRANKLIN'S WEB SITE: PRIVACY AND CURIOSITY FROM PLYMOUTH ROCK TO THE INTERNET 8–10 (2000).

[26] *See id.* at 19–20 (discussing the practice of bed-sharing primarily due to the lack of beds and for warmth).

[27] *See, e.g.*, STEPHEN MEYER III, THE FIVE DOLLAR DAY: LABOR MANAGEMENT AND SOCIAL CONTROL IN THE FORD MOTOR COMPANY 1908–1921, at 124–26 (1981); *see also* Samuel M. Levin, *Ford Profit Sharing, 1914–1920*, 6 PERSONNEL J. 75, 78 (1927).

[28] *See* SMITH, *supra* note 25, at 76; *see also* Thomas H. O'Connor, *The Right to Privacy in Historical Perspective*, 53 MASS. L. Q. 101, 104–05 (1968) (asserting that, particularly as a result of the Louisiana Purchase, "the solitary isolation of the explorers, the pioneers, and the settlers of the West was so absolute that privacy was assured by the very physical dimensions which circumscribed the frontier").

[29] *See, e.g.*, DAVID J. SEIPP, THE RIGHT TO PRIVACY IN AMERICAN HISTORY 4 (1978) (citing Commonwealth v. Lovett, 6 PA. L.J. 226, 226–28 (1847)).

[30] Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193, 205 (1891).

from the house-tops.'"[31] The latter half of the nineteenth century witnessed a change in society fueled by technological advancements, including the instant camera,[32] which itself helped fuel a profusion of newspapers and magazines satisfying an insatiable demand for gossip and intimate portrayals.[33]

[10] Soon, "[a]cceptance of the right to privacy ha[d] grown with the increasing capability of the mass media and electronic devices with their capacity to destroy an individual's anonymity, intrude upon his most intimate activities, and expose his most personal characteristics to public gaze."[34] Now, in the twenty-first century, almost all aspects of modern

[31] *Id.* at 195.

[32] *See, e.g.*, Robert E. Mensel, *"Kodakers Lying in Wait": Amateur Photography and the Right to Privacy in New York, 1885–1915*, 43 AM. Q. 24, 25 (1991) (discussing the impact of unprecedented technological change upon the public's psyche).

[33] *See* Warren & Brandeis, *supra* note 30, at 196 ("The press is overstepping in every direction the obvious bounds of propriety and of decency. Gossip is no longer the resource of the idle and of the vicious, but has become a trade, which is pursued with industry as well as effrontery. To satisfy a prurient taste the details of sexual relations are spread broadcast in the columns of the daily papers. To occupy the indolent, column upon column is filled with idle gossip, which can only be procured by intrusion upon the domestic circle."); *see also* Mensel, *supra* note 32, at 25 (arguing that "amateur photographers played an important role in provoking outrage among editorial commentators, judges, and legislators which eventually helped lead to the recognition of the right to privacy").

[34] Briscoe v. Reader's Digest Ass'n, 483 P.2d 34, 37 (Cal. 1971). In *Briscoe*, the California Supreme Court considered whether publication of a criminal's past involvement in a crime could constitute an invasion of privacy if the incident was no longer newsworthy. *See id.* at 43–44. *Briscoe* was overruled by *Gates v. Discovery Communications, Inc.*, in which the California Supreme Court ruled, under similar facts, that no invasion of privacy could occur when the broadcast in question relied on public records. *See* Gates v. Discovery Commc'ns, Inc., 101 P.3d 552, 559–62 (Cal. 2004). "[A]n invasion of privacy claim based on allegations of harm caused by a media defendant's publication of facts obtained from public official records of a criminal proceeding is barred by the First Amendment to the United States Constitution." *Id.* at 562.

life are digitally recorded, stored, and analyzed—the collection of information about us is ubiquitous.[35] Today, because of social media, mobile devices, surveillance devices, and networked sensors, individuals constantly emit information—whether they know it or not—that can be used or misused in a variety of ways.[36]

[11] Most of the information we "emit" is digital—such as e-mail and text messages, mouse clicks and keystrokes, phone numbers dialed and calls received, and GPS location data—which can suffer from over-collection and data fusion. "Over-collection occurs when an engineering design intentionally, and sometimes clandestinely, collects information

[35] *See* BIG DATA AND PRIVACY, *supra* note 7, at ix ("The ubiquity of computing and electronic communication technologies has led to the exponential growth of data from both digital and analog sources."); *see also* Diane Cardwell, *At Newark Airport, the Lights Are On, and They're Watching You*, N.Y. TIMES, Feb. 18, 2014, at A1, *available at* http://www.nytimes.com/2014/02/18/business/at-newark-airport-the-lights-are-on-and-theyre-watching-you.html, *archived at* http://perma.cc/TD3Y-FN7G ("Using an array of sensors and eight video cameras around the [Newark Airport] terminal, the light fixtures are part of a new wireless network that collects and feeds data into software that can spot long lines, recognize license plates and even identify suspicious activity, sending alerts to the appropriate staff."); Robert Faturechi, *Use of License Plate Photo Databases Is Raising Privacy Concerns*, L.A. TIMES (May 18, 2014, 9:29 PM), http://www.latimes.com/business/la-fi-law-enforcement-contractors-20140518-story.html, *archived at* http://perma.cc/A64W-2D3Z (reporting that the technology used in Los Angeles automatically captures digital images of license plates tagged with time and location, which is then transmitted to searchable databases used by police officers to track past whereabouts of drivers); Steve Lohr, *Big Data: Rise of the Machines*, N.Y. TIMES, Jan. 7, 2013, http://bits.blogs.nytimes.com/2012/12/31/big-data-rise-of-the-machines, *archived at* http://perma.cc/37SC-9H26 ("The ubiquity of sensors is new. . . . The sensors make it possible to get data we never had before." (internal quotation marks omitted)); Zach Church, *Google's Schmidt: 'Global Mind' Offers New Opportunities*, MIT NEWS (Nov. 15, 2011), http://web.mit.edu/newsoffice/2011/schmidt-event-1115.html, *archived at* http://perma.cc/3AKD-FX7P ("Technology is not really about hardware and software any more . . . . It's really about the mining and use of this enormous [volume of] data [in order to] make the world a better place." (internal quotation marks omitted)).

[36] *See* BIG DATA AND PRIVACY, *supra* note 7, at 19.

unrelated to its stated purpose."[37] For example, does your smart phone camera record your facial expressions while it records your keystrokes when you type a text message?[38] In March 2014, the FTC filed a complaint against the maker of the "Brightest Flashlight App," a popular Google Android app that would activate all the lights on a mobile device, while also transmitting the device's geolocation to third parties, including advertising networks.[39] Data fusion occurs when data collected from different sources for different reasons are brought together, resulting in data-rich profiles and new ways of tracking.[40] "[T]he privacy challenges from data fusion do not lie in the individual data streams . . . . Rather, the privacy challenges are emergent properties of our increasing ability to bring into analytical juxtaposition large, diverse data sets and to process them with new kinds of mathematical algorithms."[41]

---

[37] *Id.* at 21.

[38] *See, e.g.*, BIG DATA AND PRIVACY, *supra* note 7, at 21; *see also* Raffi Khatchadourian, *We Know How You Feel*, NEW YORKER (Jan. 19, 2015), http://www.newyorker.com/magazine/2015/01/19/know-feel, *archived at* http://perma.cc/8PMS-K3T8 (reporting on technology that can identify emotions in real time based on facial expressions: "The process requires machine learning, in which computers find patterns in large tranches of data, and then use those patterns to interpret new data.").

[39] *See* Complaint at 2, *In re* Goldenshores Tech., L.L.C., No. C-4446 (F.T.C. Mar. 31, 2014), *available at* http://www.ftc.gov/system/files/documents/cases/140409goldenshorescmpt.pdf, *archived at* http://perma.cc/9HWU-NG6G. The FTC entered a settlement with Goldenshores Technologies that requires it to delete personal information collected from consumers as well as provide a just-in-time disclosure that fully informs consumers when, how, and why their geolocation information is being collected, used and shared, and requires defendants to obtain consumers' affirmative express consent before doing so. *See* Decision and Order at 4, *In re* Goldenshores Tech., L.L.C., No. C-4446 (F.T.C. Mar. 31, 2014), *available at* http://www.ftc.gov/system/files/documents/cases/140409goldenshoresdo.pdf, *archived at* http://perma.cc/Y6FW-C2M4.

[40] *See* BIG DATA AND PRIVACY, *supra* note 7, at 21.

[41] *Id.*

[12] This is one way in which predictive analytics contribute to online tracking.[42] One does not even have to shop online to be targeted by predictive analytics. Perhaps the most famous—and chilling—example comes from Target Corporation's use of analytics to predict its shoppers' future buying habits. Target—like all other retailers—understands that many consumers' buying habits are ingrained and difficult to change.[43] But, one particular moment when buying habits can change significantly is the birth of a child. Most marketers are reactive, and send coupons and advertisements after the birth of the child based on public birth records. Target sought to be proactive—predicting when shoppers, based on buying habits,[44] are pregnant.[45] Unfortunately, Target's analytics were so

[42] *See, e.g.*, Julie Brill, *Competition and Consumer Protection: Strange Bedfellows or Best Friends?*, 10 ANTITRUST SOURCE, Dec. 2010, at 7, *available at* http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/Dec10_Brill12_21f.authcheckdam.pdf, *archived at* http://perma.cc/45VS-UX28 ("In recent years, advances in computer technology have made it possible for detailed information about consumers to be stored, sold, shared, aggregated, and used more easily and cheaply than ever, in ways not feasible, or even conceivable, before. These advances in technology have allowed online companies to engage in targeted advertising, a practice that has many important benefits. . . . Yet serious privacy concerns arise when companies can easily collect, combine, and use so much information from consumers.").

[43] *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG., Feb. 19, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html, *archived at* http://perma.cc/Q6RQ-5Q67.

[44] *See id.* Whenever possible, Target assigns each shopper a unique code that tracks every purchase and which is linked to individual demographic information, such as age, marriage status, neighborhood, estimated salary, credit cards used, and Web sites visited. *See id.*

> Target can buy data about your ethnicity, job history, the magazines you read, if you've ever declared bankruptcy or got divorced, the year you bought (or lost) your house, where you went to college, what kinds of topics you talk about online, whether you prefer certain brands of coffee, paper towels, cereal or applesauce, your political leanings, reading habits, charitable giving and the number of cars you own.

Id.

good it informed a father of his daughter's pregnancy before he even knew about it.[46] As the Target incident illustrates, predictive analytics can create a risk of revealing intimate personal information before it becomes publicly available,[47] even when the original data is non-personally identifiable.[48]

[45] *See id.* Rather than be among the multiple marketers sending materials post-birth, Target wanted to target (pun intended) women in their second trimester, which is when most expectant mothers begin buying all new and different items, such as prenatal vitamins and maternity clothing. *See id.*

[46] *See* Duhigg, *supra* note 43 (relating how an irate father complained to Target that it was encouraging his teenage daughter to get pregnant by sending her coupons for maternity and baby clothes and cribs, only to find out later that his daughter was indeed pregnant). *But see* Tim Harford, *Big Data: Are We Making a Big Mistake?*, FIN. TIMES (Mar. 29, 2014, 11:38 AM), http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz3TwAfkZFe, *archived at* http://perma.cc/2JAW-ZFTK (arguing that pregnant women receive pregnancy-related coupons because everyone on Target's mailing list receives such coupons; suggesting further that Target mixes pregnancy-related coupons with other unrelated coupons not to avoid upsetting pregnant women who would receive only pregnancy-related coupons but because the coupons will be sent to women who are not pregnant); *cf.* Jessica Goldstein, *Meet the Woman Who Did Everything in Her Power to Hide Her Pregnancy from Big Data*, THINKPROGRESS (Apr. 29, 2014, 11:26 AM), http://thinkprogress.org/culture/2014/04/29/3432050/can-you-hide-from-big-data/, *archived at* http://perma.cc/R7Q7-KEQJ (reporting on the efforts of Jane Vertesi, Princeton University Assistant Professor of Sociology, to prevent "big data" from finding out she was pregnant, and who found hiding from "big data" extremely inconvenient and expensive, besides nearly impossible).

[47] *See* BIG DATA AND ANALYTICS, *supra* note 4, at 2 ("[T]he power of analytics, rich data stores and the insights they can yield raise risks to privacy."); Tene & Polonetsky, *supra* note 8, at 253–54 ("It is one thing to recommend for a customer books, music or movies she might be interested in based on her previous purchases; it is quite another thing to identify when she is pregnant before her closest family knows.").

[48] *See* BIG DATA AND ANALYTICS, *supra* note 4, at 2; *see also* Justin Brickell & Vitaly Shmatikov, *The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing* (Aug. 2008) (presented at 14th Annual ACM SIGKDD Conference on Knowledge, Discovery and Data Mining in Las Vegas, Nev), *available at* http://www.cise.ufl.edu/~nemo/anonymity/papers/jlbrick_kdd2008.pdf, *archived at* http://perma.cc/Q8LA-3HNY ("[E]ven modest privacy gains require almost complete destruction of the data-mining utility."). *See generally* Paul Ohm, *Broken Promises of*

> Application of analytics to big data does not conform well to traditional legal approaches because big data does not result from one-on-one interaction between the data controller and the individual. Big data instead pulls in information from disparate sources. Its value derives not only from its volume, but also from its varied and expansive scope—big data brings together an enormous pool of information that initially may seem unrelated.[49]

A. The Public/Private Dichotomy and the Third-Party Doctrine

[13] The principal privacy conundrum posed by predictive analytics is that data mining relies to a large extent on "public" information—it derives from transactions and social interactions that are often generally observable. "A matter that is already public or that has previously become part of the public domain is not private."[50] While total secrecy is not required—information disclosed to a few people may remain private[51]—if even only a few people actually see the information, privacy can be lost if the *potential* audience is large.[52] W.A. Parent expressly excludes information in the public domain from his definition of privacy, considering it a "glaring paradox."[53] Lior Jacob Strahilevitz considers the

Privacy: Responding to the Surprising Failure of Anonymization, 57 UCLA L. REV. 1701, 1706–07 (2010) (demonstrating that cross-linking supposedly anonymous data among databases can "de-anonymize" the data).

[49] BIG DATA AND ANALYTICS, *supra* note 4, at 11.

[50] Moreno v. Hanford Sentinel, Inc., 91 Cal. Rptr. 3d 858, 862 (Cal. Ct. App. 2009).

[51] *See id.* at 863 (citing M.G. v. Time Warner, Inc., 107 Cal. Rptr. 2d 504, 511 (Cal. Ct. App. 2001)).

[52] *See id.* at 863 (holding that posting information to MySpace opened it to the public at large, even if it was removed a few days later and was seen by only a few people).

[53] W.A. Parent, *Privacy, Morality, and the Law*, 12 PHIL. & PUB. AFF. 269, 271 (1983).

boundary between public and private "*the* fundamental, first-principles question in privacy law."[54] This public/private dichotomy is reflected in a number of court decisions: "objects, activities, or statements that [one] exposes to the 'plain view' of outsiders are not 'protected'";[55] "whatever the public may see from a public place cannot be private";[56] and watching an appellee and videotaping his activities while he was outside his home, in his front yard, where he was exposed to public view was not an actionable invasion of privacy.[57] Indeed, one court has gone so far as to hold there was no reasonable expectation of privacy where a woman was recorded by a secretly installed camera while changing clothes in an office area, despite locking the door, because others had a key to the office and could have walked in at any moment.[58]

[14] Closely related to the public/private dichotomy is the so-called "third-party doctrine," which provides that private information disclosed to a third party can lose its privacy protection. The doctrine originates in Fourth Amendment jurisprudence, particularly in the following cases: *On Lee v. United States*, where the Supreme Court held there was no Fourth Amendment protection in a confidential conversation recorded by an informant;[59] *United States v. Miller*, where the Supreme Court held that

[54] Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919, 920 (2005) (emphasis added).

[55] Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

[56] N.O.C., Inc. v. Schaefer, 484 A.2d 729, 732 n.1 (N.J. Super. Ct. Law Div. 1984).

[57] *See* I.C.U. Investigations, Inc. v. Jones, 780 So. 2d 685, 689–90 (Ala. 2000).

[58] *See* Nelson v. Salem State Coll., 845 N.E.2d 338, 346–47 (Mass. 2006).

[59] *See* On Lee v. United States, 343 U.S. 747, 753–54 (1952) ("Petitioner was talking confidentially and indiscreetly with one he trusted, and he was overheard. This was due to aid from a transmitter and receiver, to be sure, but with the same effect on his privacy as if agent Lee had been eavesdropping outside an open window. The use of bifocals, field glasses or the telescope to magnify the object of a witness' vision is not a forbidden search or seizure, even if they focus without his knowledge or consent upon what one supposes to be private indiscretions."). *But cf.* Kyllo v. United States, 533 U.S. 27, 34–

the Fourth Amendment does not require the government to obtain a warrant to seize bank records;[60] and *Smith v. Maryland*, where the Court held that dialed telephone numbers have no constitutional protection.[61]

[15] The third-party doctrine is not without its critics.[62] One argument

35 (2001) (holding that use of thermal imaging technology constituted a Fourth Amendment search).

[60] *See* United States v. Miller, 425 U.S. 435, 443 (1976) ("[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."). *Miller*'s holding was limited by the Right to Financial Privacy Act of 1978, by—for example—requiring the Government authority to notify the bank customer of the subpoena or summons served on the financial institution as well as the nature of the law enforcement inquiry to which the subpoena or summons relates. *See* Right to Financial Privacy Act of 1978, Pub. L. No. 95-630, 92 Stat. 3697 (1978) (codified as amended at 12 U.S.C. §§ 3401–3422 (2012)).

[61] *See* Smith v. Maryland, 442 U.S. 735, 743–46 (1979) ("Although petitioner's conduct may have been calculated to keep the *contents* of his conversation private, his conduct was not and could not have been calculated to preserve the privacy of the number he dialed. Regardless of his location [i.e., whether in his home or some other location], petitioner had to convey that number to the telephone company in precisely the same way if he wished to complete his call."). *But see* Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, tit. III, 100 Stat. 1848, 1868 (codified as amended at 18 U.S.C. § 3121–23) (2012)) (requiring government authorities to obtain a court order prior to recording telephone numbers dialed).

[62] *See, e.g.*, *Smith*, 442 U.S. at 750 (Marshall, J., dissenting) ("[U]nless a person is prepared to forgo use of what for many has become a personal or professional necessity [i.e., the telephone], he cannot help but accept the risk of surveillance."); United States v. White, 401 U.S. 745, 790 (1971) (Harlan, J., dissenting) ("The interest *On Lee* fails to protect is the expectation of the ordinary citizen, who has never engaged in illegal conduct in his life, that he may carry on his private discourse freely, openly, and spontaneously without measuring his every word against the connotations it might carry when instantaneously heard by others unknown to him and unfamiliar with his situation or analyzed in a cold, formal record played days, months, or years after the conversation. Interposition of a warrant requirement is designed not to shield 'wrongdoers,' but to secure a measure of privacy and a sense of personal security throughout our society."); *cf.* Frank Rich, *Chris Rock Talks to Frank Rich about Ferguson, Cosby, and What 'Racial*

particularly germane to this article is that privacy does not require total secrecy and that exposure to a limited audience does not equate to exposure to the world at large.[63]

[16] In contrast, Orin Kerr argues the third-party doctrine prevents "savvy wrongdoers" from using "third-party services in a tactical way to

Progress' Really Means, VULTURE (Nov. 30, 2014, 9:00 PM), http://www.vulture.com/2014/11/chris-rock-frank-rich-in-conversation.html, *archived at* http://perma.cc/23Z4-SA8T ("You can't think the thoughts you want to think if you think you're being watched.").

[63] *See, e.g.*, RICHARD A. POSNER, NOT A SUICIDE PACT: THE CONSTITUTION IN A TIME OF NATIONAL EMERGENCY 140 (2006) ("Information privacy does not mean refusing to share information with everyone. . . . One must not confuse solitude with secrecy; they are distinct forms of privacy."); Sherry F. Colb, *What Is a Search? Two Conceptual Flaws in Fourth Amendment Doctrine and Some Hints of a Remedy*, 55 STAN. L. REV. 119, 122–23 (2002) (arguing that "treating exposure to a limited audience as identical to exposure to the world" fails "to recognize degrees of privacy in the Fourth Amendment context;" noting for example, that giving a neighbor keys to one's house so the neighbor can water the plants while the owner is away does not grant the neighbor permission to invite friends into the owner's house); Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1086–87 (2002) ("The [Supreme] Court's current conception of privacy is as a form of total secrecy. As conceived by the Court, an individual's hidden world should be protected. It has expressed an interest in safeguarding the intimate information that individuals carefully conceal. Privacy is about protecting the skeletons that are meticulously hidden in the closet. Since information maintained by third parties is exposed to others, it is not private, and therefore not protected by the Fourth Amendment. This conception of privacy is not responsive to life in the modern Information Age."). *But see* State v. Carle, 337 P.3d 904, 911 (Or. Ct. App. 2014) (concluding that because defendant had no privacy interest in the digital copy of the text message found on the recipient's phone, police did not conduct a "search" under the Fourth Amendment when they viewed that text message on recipient's phone); State v. Patino, 93 A.3d 40, 56 (R.I. 2014) (holding that a sender of text messages has no Fourth Amendment privacy interest in those messages stored on the recipient's phone, "[b]ecause the recipient now shares full control of whether to share or disseminate the sender's message, the sender, to be sure, no longer enjoys a reasonable expectation of privacy in the digital copy of the message contained on the recipient's device."); State v. Marcum, 319 P.3d 681, 687 (Okla. Crim. App. 2014) (holding defendant had no reasonable expectation of privacy in phone company's records of defendant's text messages from account of co-defendant).

enshroud the entirety of their crimes in zones of Fourth Amendment protection."[64] However, the Supreme Court may ultimately recognize that modern technology may finally impose a limit on the Fourth Amendment's third-party doctrine:

> [I]t may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. This 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.[65]

[17] The third-party doctrine has been applied in common law privacy cases as well. For example, the United States District Court for the District of New Jersey dismissed a plaintiff's invasion of privacy claim against her employer in relation to restricted-access Facebook posts because a person allowed to view those posts had provided them to her employer.[66] In *Sumien v. CareFlite*, the Texas Court of Appeals refused

[64] Orin S. Kerr, *The Case For The Third-Party Doctrine*, 107 MICH. L. REV. 561, 564–65 (2009) (also arguing that the third-party doctrine provides clarity by focusing on the information's knowable location rather than its unknowable history). For further elaboration and debate regarding Kerr's arguments, *see* Richard A. Epstein, Symposium, *Security Breach Notification Six Years Later: Privacy and the Third Hand: Lessons from the Common Law of Reasonable Expectations*, 24 BERKELEY TECH. L.J. 1199, 1200–02 (2009) (stating "[i]nsofar as Kerr shifts away from reasonable expectations to either assumption of risk or to consent, he cannot build an adequate foundation for Fourth Amendment Law."); Erin Murphy, *Security Breach Notification Six Years Later: The Case Against the Case for Third-Party Doctrine: A Response to Epstein and Kerr*, 24 BERKELEY TECH. L.J. 1239, 1241–44 (2009) (stating that it "is not clear that third-party participation makes policing all that harder or easier."); *see also* Orin S. Kerr, Symposium, *Security Breach Notification Six Years Later: Defending The Third-Party Doctrine: A Response to Epstein and Murphy*, 24 BERKELEY TECH. L.J. 1229, 1230–36 (2009) (addressing the concerns raised by Epstein and Murphy).

[65] United States v. Jones, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring) (citations omitted).

[66] *See* Ehling v. Monmouth-Ocean Hosp. Serv. Corp., 961 F. Supp. 2d 659, 673–74 (D.N.J. 2013) ("[T]he evidence shows that Defendants were the passive recipients of

to recognize a right of privacy in Facebook posts viewed by a friend-of-a-friend.[67]

[18] Big data and its associated predictive analytics extend the privacy concern beyond discrete events and transactions—a concern perhaps best described by a fictional professor describing his forthcoming book *The Defeat of Privacy*:

> It's about the fact that there are no more private selves, no more private corners in society, no more private properties, no more private acts. . . . Mankind is making everything open and accessible. . . . There are no concealments any longer, no mysterious dark places of the soul. We're all right there in front of the entire audience of the universe, in a state of exposure. We're all nude and available.[68]

information that they did not seek out or ask for. Plaintiff voluntarily gave information to her Facebook friend, and her Facebook friend voluntarily gave that information to someone else. This may have been a violation of trust, but it was not a violation of privacy.").

[67] *See* Sumien v. CareFlite, No. 02–12–00039–CV, 2012 Tex. App. LEXIS 5331, at *6–7 (Tex. Ct. App. July 5, 2012). *But cf.* Pure Power Boot Camp, Inc. v. Warrior Fitness Boot Camp, LLC, 587 F. Supp. 2d 548, 561 (S.D.N.Y. 2008) ("There is no sound basis to argue that Fell [plaintiff's former employee], by inadvertently leaving his Hotmail password accessible, was thereby authorizing access to all of his Hotmail e-mails, no less the e-mails in his two other accounts. If he had left a key to his house on the front desk at [Pure Power Boot Camp], one could not reasonably argue that he was giving consent to whoever found the key, to use it to enter his house and rummage through his belongings. And, to take the analogy a step further, had the person rummaging through the belongings in Fell's house found the key to Fell's country house, could that be taken as authorization to search his country house. We think not."); Colb, *supra* note 63, at 122–23.

[68] MALCOLM BRADBURY, THE HISTORY MAN 73 (1976) (internal quotation marks omitted). When the person to whom the fictional professor is describing his book asks, "You mean there isn't a me anymore?" the professor responds, "You're there, you're present . . . but you happen to be a conjunction of known variables . . . ." *Id.* (internal quotation marks omitted).

In light of ever-developing technologies that provide ubiquitous tracking and data collection, perhaps, as Justice Sotomayor intimated, it is time to reexamine the public/private dichotomy and third-party doctrine. This reexamination is currently taking place in Fourth Amendment cases and can easily be applied to common law privacy.

B. "Public" Data and the Mosaic Theory

[19] In 2001, Daniel Solove identified the risk to privacy imposed by data analytics: "It is ever more possible to create an electronic collage that covers much of a person's life—a life captured in records, a digital biography composed in the collective computer networks of the world."[69] Herman Tavani succinctly summarizes the fundamental conundrum between data mining (and implicitly predictive analytics) and privacy:

> Unlike personal data that reside in explicit records in databases, information acquired about persons via data mining is often derived from implicit patterns in the data. The patterns can suggest "new" facts, relationships, or associations about a person, placing that person in a "newly discovered" category or group. Also, because most personal data collected and used in data mining applications is considered neither confidential nor intimate in nature, there is a tendency to presume that such data must by default be *public* data. And unlike the personal data that are often exchanged between or across two or more databases in traditional database retrieval processes, in the data mining process personal data are often manipulated within a single database, and typically within a large *data warehouse*.[70]

[69] Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1394 (2001).

[70] TAVANI, *supra* note 3, at 151.

[20] Derived from government surveillance cases,[71] the "mosaic theory" recognizes that continual surveillance of a suspect's public movements "reveals far more than the individual movements [the whole] comprises."[72] The current leading Fourth Amendment case espousing the mosaic theory is *United States v. Maynard*, in which law enforcement agents tracked a suspect continuously for a month using a GPS device attached to his car.[73] The D.C. Circuit Court of Appeals concluded that the GPS tracking constituted a search within the meaning of the Fourth Amendment, and therefore required a warrant.[74] While the U.S. Supreme

[71] *See, e.g.*, Halkin v. Helms, 598 F.2d 1, 8 (D.C. Cir. 1978) ("It requires little reflection to understand that the business of foreign intelligence gathering in this age of computer technology is more akin to the construction of a mosaic than it is to the management of a cloak and dagger affair. Thousands of bits and pieces of seemingly innocuous information can be analyzed and fitted into place to reveal with startling clarity how the unseen whole must operate."); *see also* Halperin v. C.I.A., 629 F.2d 144, 150 (D.C. Cir. 1980) ("We must take into account . . . that each individual piece of intelligence information, much like a piece of jigsaw puzzle, may aid in piecing together other bits of information even when the individual piece is not of obvious importance in itself.")

[72] United States v. Maynard, 615 F.3d 544, 561–62 (D.C. Cir. 2010), *aff'd sub nom.* United States v. Jones, 132 S. Ct. 945 (2012)*; see also* Smith v. Maryland, 442 U.S. 735, 748 (1979) (Stewart, J., dissenting) (arguing that a list of telephone numbers dialed could "reveal the most intimate details of a person's life."); United States v. Pineda-Moreno, 617 F.3d 1120, 1120, 1125 (9th Cir. 2010) (denying petition for rehearing en banc) (Kozinski, C.J., dissenting) ("By tracking and recording the movements of millions of individuals the government can use computers to detect patterns and develop suspicions. It can also learn a great deal about us because where we go says much about who we are. Are Winston and Julia's cell phones together near a hotel a bit too often? Was Syme's OnStar near an STD clinic? Were Jones, Aaronson and Rutherford at that protest outside the White House? The FBI need no longer deploy agents to infiltrate groups it considers subversive; it can figure out where the groups hold meetings and ask the phone company for a list of cell phones near those locations."); *In re* United States Order Authorizing the Release of Historical Cell-Site Info., 736 F. Supp. 2d 578, 584–86, 589–95 (E.D.N.Y 2010) (applying *Maynard* analysis in to order to determine the difference between historical and perspective tracking; surveillance and disclosure; precision capabilities of CSI and GPS; and vehicle and telephone trafficking).

[73] *See Maynard*, 615 F.3d at 549.

[74] *See id.* at 555–56, 563 ("Society recognizes [the suspect's] expectation of privacy in his movements over the course of a month as reasonable, and the use of the GPS device

Court affirmed *Maynard*, it did not do so under the mosaic theory—it instead held that the agents needed a warrant because they physically trespassed when they placed the GPS on the suspect's car.[75] However, in his concurrence with the judgment, Justice Alito expressly stated that he would "analyze the question presented in this case by asking whether respondent's reasonable expectations of privacy were violated by the long-term monitoring of the movements of the vehicle he drove."[76] Justice Alito suggested that the majority's "reasoning largely disregards what is really important (the *use* of a GPS for the purpose of long-term tracking) and instead attaches great significance to something that most would view as relatively minor"—the attaching to the bottom of a car the GPS device itself.[77] And as noted earlier, Justice Sotomayor expressed her opinion "it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties."[78]

---

to monitor those movements defeated that reasonable expectation. . . . [P]rolonged GPS monitoring reveals an intimate picture of the subject's life that he expects no one to have . . . ."); *see also* Benjamin Zhu, Note, *A Traditional Tort for a Modern Threat: Applying Intrusion upon Seclusion to Dataveillance Observations*, 89 N.Y.U. L. REV. 2381, 2405–06 (2014) ("*Maynard* implicitly recognizes that an individual may have a privacy interest in the inferences and new information that may be drawn from a collection of data, even if each piece of data had been disclosed to the public.").

[75] *See Jones*, 132 S. Ct. at 949, 950 n.3 ("Where, as here, the Government obtains information by physically intruding on a constitutionally protected area . . . a search [requiring a warrant] has undoubtedly occurred."). Indeed, the majority openly skirted what many commentators and Court observers considered to be the key issue in the case: "It may be that achieving the same result [i.e., continuous surveillance for a 4-week period] through electronic means, without an accompanying trespass, is an unconstitutional invasion of privacy, but the present case does not require us to answer that question." *Id.* at 954.

[76] *Id.* at 958 (Alito, J., concurring).

[77] *Id.* at 961. Justice Alito noted that the basis for the Court's holding in *Jones* would be inapplicable once all cars were fitted with GPS devices. *See id.*

[78] *See supra* text accompanying note 65.

[21] One can also perhaps glean some insight into the Supreme Court's concern over Fourth Amendment privacy and evolving technology in *Riley v. California*, in which the Court held that warrantless searches of cell phones incident to arrest violated the Fourth Amendment.[79] As a general matter, police officers do not need a warrant to search a person and the immediately surrounding area incident to an arrest in order to secure the officers' safety and prevent the destruction evidence.[80] In *Riley*, the Supreme Court concluded that neither risk exists when the search is of digital data.[81] Ultimately, "when 'privacy-related concerns are weighty enough' a 'search may require a warrant, notwithstanding the diminished expectations of privacy of the arrestee.'"[82]

[22] Importantly, at least for the arguments made in this article, Justice Robert's opinion recognizes that collected data can be different:[83]

> The storage capacity of cell phones has several interrelated consequences for privacy. First, a cell phone collects in one place many distinct types of information—an address, a note, a prescription, a bank statement, a video—*that reveal much more in combination than any isolated record*. Second, a cell phone's capacity allows even just one type of information to convey far more than previously possible. *The sum of an individual's private life can be reconstructed*

[79] *See* Riley v. California, 134 S. Ct. 2473, 2480, 2493 (2014). Seven Justices joined Chief Justice Robert's opinion, with Justice Alito concurring in part and concurring in the judgment. *See id.* at 2480. *Riley* consolidated two cell phone search cases, one involving a smart phone and one involving a flip phone. *See id.* at 2480–81.

[80] *See id.*, 134 S. Ct. at 2483 (citing Chimel v. Cal., 395 U.S. 752, 762–63 (1969)).

[81] *See id.* at 2484–85.

[82] *Id.* at 2488 (quoting Maryland v. King, 133 S. Ct. 1958, 1979 (2013)).

[83] *See Riley*, 134 S. Ct. at 2489 ("Cell phones differ in both a quantitative and a qualitative sense from other objects that might be kept on an arrestee's person.").

> *through a thousand photographs labeled with dates, locations, and descriptions*; the same cannot be said of a photograph or two of loved ones tucked into a wallet. Third, the data on a phone can date back to the purchase of the phone, or even earlier. A person might carry in his pocket a slip of paper reminding him to call Mr. Jones; he would not carry a record of all his communications with Mr. Jones for the past several months, as would routinely be kept on a phone.[84]

[23] Analogous to ubiquitous monitoring, Chief Justice Roberts additionally noted the pervasiveness of data collected through cell phones: "Today . . . it is no exaggeration to say that many of the more than 90% of American adults who own a cell phone keep on their person a digital record of nearly every aspect of their lives—from the mundane to the intimate."[85] Chief Justice Roberts noted also the qualitative difference in stored data (in this case, on a cell phone):

> An Internet search and browsing history . . . could reveal an individual's private interests or concerns . . . . Data on a cell phone can also reveal where a person has been. Historic location information . . . can reconstruct someone's specific movements down to the minute, not only around town but also within a particular building.[86]

[84] *Id.* (emphasis added).

[85] *Id.* at 2490 (2014); *see also* United States v. Zavala, 541 F.3d 562, 577 (5th Cir. 2008) (holding suspect had privacy right in cell phone data; likening a cell phone to a personal computer carried on one's person; noting "cell phones contain a wealth of private information, including e[-]mails, text messages, call histories, address books, and subscriber numbers."); State v. Marcum, 319 P.3d 681, 685–87 (Okla. Crim. App. 2014) (reviewing various state courts which have recognized privacy interests in cell phone data).

[86] *Riley*, 134 S. Ct. at 2490.

[24] Based on *Riley*'s discussion of the expanding quantitative and qualitative aspects of stored data, it is logical to surmise that if the Supreme Court had addressed the substantive privacy issues in *Jones*, it very likely might have adopted the mosaic theory–particularly in light of Justice Alito's and Justice Sotomayor's concurrences. [87] One distinguishing factor in "mosaic" cases is the length of surveillance. "[R]elatively short-term monitoring of a person's movements on public streets accords with expectations of privacy that our society has recognized as reasonable. But the use of longer term GPS monitoring in investigations of most offenses impinges on expectations of privacy."[88] Indeed this was one of Justice Scalia's objections to applying the theory—"it remains unexplained why a 4-week investigation is 'surely' too long."[89] This may be an issue within Fourth Amendment jurisprudence, but it should not be in the context of private-party data tracking and analysis—it is now ubiquitous and fundamentally unavoidable.

[25] Within the context of private-party data tracking and analysis, the mosaic theory is less about length than extremity. New York courts in particular have recognized a common law privacy invasion resulting from overly zealous surveillance of "public" conduct.[90] For example, in *Nader v. General Motors Corporation*, in which General Motors had hired

[87] *Cf.* Note, *Data Mining, Dog Sniffs, and The Fourth Amendment*, 128 HARV. L. REV. 691, 701 (Dec. 2014) (concluding that while cases such as *Jones* and *Riley* "suggest that the Court is aware that modern surveillance technologies represent a problem for traditional Fourth Amendment doctrine, but is still casting about for a solution that might prove workable in the context of data mining.").

[88] United States v. Jones, 132 S. Ct. 945, 964 (Alito, J., concurring) (citations omitted) ("For… [most] offenses, society's expectation has been that law enforcement agents and others would not—and indeed, in the main, simply could not secretly monitor and catalogue every single movement of an individual's car for a very long period.").

[89] *Id.* at 954.

[90] *See generally* Nader v. Gen. Motors Corp., 255 N.E.2d 765, 771 (N.Y. 1970) (discussing how surveillance in public places may be so "overzealous" that it could be rendered as actionable).

private investigators to follow Ralph Nader, a critic of General Motors, and interview his acquaintances, the New York Court of Appeals concluded that surveillance of public activities could rise to the level of an invasion of privacy.[91] "[I]t is manifest that the mere observation of the plaintiff in a public place does not amount to an invasion of his privacy. But, under certain circumstances, surveillance may be so 'overzealous' as to render it actionable."[92] Judge Breitel elaborated: "Although acts performed in 'public', especially if taken singly or in small numbers, may not be confidential, at least arguably a right to privacy may nevertheless be invaded through extensive or exhaustive monitoring and cataloguing of acts *normally disconnected* and anonymous."[93] Similarly, in *Galella v. Onassis*, a photographer who had stalked former First Lady Jacqueline Onassis to such an extent that he could comment "at considerable length on her personality, her shopping tastes and habits, and her preferences for entertainment," had invaded Onassis's privacy.[94]

[26] The Supreme Court has expanded the concept that some public information can be private. Recognizing that "both the common law and the literal understandings of privacy encompass the individual's control of

[91] *Id.* at 770–71 (applying District of Columbia law).

[92] *Id.* (citing Pinkerton Nat'l Detective Agency, Inc. v. Stevens, 132 S.E.2d 119 (Ga. Ct. App. 1963) (holding that allegations that insurer had detective agency constantly shadow woman after she filed personal injury action against the insurer in a manner calculated to frighten her and give her neighbors the impression that she was engaged in some wrongful activity were sufficient for a claim of invasion of privacy)).

[93] Nader, 255 N.E.2d at 770, 772 (Breitel, J., concurring) (emphasis added) (noting that the New York Court of Appeals did rule, though, that the investigators' interviewing Mr. Nader's acquaintances did not violate his privacy and that "[i]nformation about the plaintiff which was already known to others could hardly be regarded as private to the plaintiff").

[94] *See* Galella v. Onassis, 353 F. Supp. 196, 228 (S.D.N.Y. 1972) (applying New York law) ("The surveillance, close-shadowing and monitoring were clearly "overzealous" and therefore actionable."), *aff'd* in relevant part, 487 F.2d 986 (2d Cir. 1973).

information concerning his or her person"[95] in *Department of Justice v. Reporters Committee for Freedom of the Press*, Justice Stevens focused on "whether the compilation of otherwise hard-to-obtain information alters the privacy interest implicated by disclosure of that information."[96] Recognizing that complete computerized dossiers are now available at one's fingertips,[97] Justice Stevens concluded that the Freedom of Information Act's exemptions from disclosure recognize "the power of compilations to affect personal privacy that outstrips the combined power of the bits of information contained within."[98] Fundamentally, the Supreme Court has "recognized the privacy interest inherent in the nondisclosure of certain information even where the information may have been at one time public."[99]

[27] Extensive data mining and the use of predictive analytics appear to fit squarely within the concerns expressed by advocates of the mosaic theory, as well as those who have expressed similar concerns regarding data agglomeration. Citizens should not have to fear that personal and intimate details of their lives will be revealed through the collection, storage, and analysis of virtually all of the mundane acts of life.[100] For all

[95] Dep't of Justice v. Reporters Comm. for Freedom of Press, 489 U.S. 749, 763 (1989).

[96] *Id.* at 764 (1989) (noting "there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information") (applying the Freedom of Information Act, Pub. L. No. 89-554, 80 Stat. 383 (1966) (codified as amended at 5 U.S.C. § 552 (2012))).

[97] *See* Kenneth L. Karst, *The Files: Legal Controls Over The Accuracy and Accessibility of Stored Personal Data*, 31 LAW & CONTEMP. PROBS. 342, 343 (1966).

[98] *Reporters Comm. for Freedom of Press*, 489 U.S. at 765 (applying 5 U.S.C. § 552(b)(3)).

[99] *Id.* at 767; *see also* Whalen v. Roe, 429 U.S. 589, 605 (1977) ("We are not unaware of the threat to privacy implicit in the accumulation of vast amounts of personal information in computerized data banks or other massive government files.").

[100] Indeed, one court has applied *Riley*. *See* Bakhit v. Safety Marking, Inc., No. 3:13CV1049 (JCH) 2014 U.S. Dist. LEXIS 86761, at *8–9 (D. Conn. June 26, 2014); *see*

practical purposes, it is now impossible to avoid being tracked:

> Experience has shown that it is *possible*, but it's really not easy, and it comes with a lot of sacrifices. And it requires some technical skill. So to that end, it's my concern about the opt-out idea. I don't actually think it's feasible for everyone to do this. I don't think that's the answer. I don't think that's the simple answer to the big data problem: that you can just turn this stuff off, that you cannot do the things that you clearly need to do for your daily life. But I really want to emphasize, I did this [avoiding tracking] as an experiment to see what it would take, to see what these systems were demanding of us that we'd forgotten about, and how it is that they worked. And so I don't expect people to do this. In fact, I wouldn't recommend it.[101]

[28] No one should be forced into the experimental "dilemma" attempted by Professor Vertesi of completely rejecting all aspects of modern life—from social communications to shopping—to avoid constant commercial surveillance.[102] But in the workplace, we have even less control over the degree of monitoring that is taking place.

---

also supra text accompanying notes 79–86 (discussing the holding of *Riley*). In *Bakhit v. Safety Marking, Inc*., plaintiffs sought to view the contents of the cell phones of ten Safety Marking employees, seeking evidence of racially offensive text messages and images. Bakhit, 2014 LEXIS 86761 at *3–4. Magistrate Judge Fitzsimmons denied (without prejudice) the plaintiff's discovery request based, in part, on *Riley*, to protect the individual employees' privacy interests in the contents of their cell phones. *See id.* at *9–10

[101] Goldstein, *supra* note 46; *see also* Chris Jay Hoofnagle et al., Symposium, *Privacy and Accountability in the 21st Century: Behavioral Advertising: The Offer You Cannot Refuse*, 6 HARV. L. & POL'Y REV. 273, 273 (2012) ("[A]vertisers use new, relatively unknown technologies to track people, specifically because consumers have not heard of these techniques. Furthermore, these technologies obviate choice mechanisms that consumers exercise.").

[102] *See* Goldstein, *supra* note 46.

IV. PREDICTIVE ANALYTICS IN THE WORKPLACE

[29] In the late nineteenth and early twentieth centuries, Frederick Taylor and his scientific management techniques sought to obtain maximum efficiency in industrial work—the one best way.[103] Under Taylor's methodology, the principal tools of measurement and observation were the manager with a stopwatch.[104] To the workers, the stopwatch "was a hideous invasion of privacy, an oppressive all-seeing eye that peered into their work lives, ripping at their dignity."[105]

[30] Fast forward to the twenty-first century. Under what Simon Head describes as Computer Business Systems, the shop floor has moved into service-sector offices:

> With the coming of the networked computer with monitoring software attached, industrial regimes of quantification, targeting, and control now pervade the white-collar world: how many patients, litigants, customers with complaints, students with theses, and future home owners with mortgage applications have been processed or billed per day or week, and how many *should* be processed or billed, because the digital white-collar line is subject to speedup no less than its factory counterpart?[106]

[103] *See* Robert Kanigel, *Taylor-Made: How The World's First Efficiency Expert Refashioned Modern Life in His Own Image*, 37 SCI. 13, 18–19 (1997).

[104] *See* ROBERT KANIGEL, THE ONE BEST WAY: FREDERICK WINSLOW TAYLOR AND THE ENIGMA OF EFFICIENCY 466 (1997).

[105] *Id.*

[106] SIMON HEAD, MINDLESS: WHY SMARTER MACHINES ARE MAKING DUMBER HUMANS 5 (2014); *see also* Monika Bauerlein & Clara Jeffery, *All Work and No Pay: The Great Speedup*, MOTHER JONES, July-Aug. 2011, at 18, 19 (describing the modern phenomenon of speedup whereby workers are pressured to work more hours with no additional pay), *available at* http://www.motherjones.com/politics/2011/06/speed-up-american-workers-long-hours, *archived at* http://perma.cc/G7MJ-PXTF.

A. Predictive Analytics: Surveillance on Steroids?

[31] One commentator from Frederick Taylor's era suggested that any system that schedules every movement so that a worker "is simply one of the gears in the operation of the machine" would not secure the best results in the long run.[107] Adding big data and analytics to the mix may only compound the problems:

> Big Data also holds out the promise of, for instance, total supervision in the workplace. Lest perfect surveillance of employees sound alarming, this new field is given the blandly technocratic name of "workforce science". Every phone call, email and even mouse-click of an employee can be stored and analysed to guide management in making decisions.
>
> So "workforce science" is a scaled-up and automated version of the "scientific management" promoted by Frederick Winslow Taylor in his highly influential 1911 book, *The Principles of Scientific Management*, which recounted how he performed time-and-motion studies on labourers in order to get more work out of them. It has since been alleged that Winslow fiddled the data, but that didn't stop him becoming an eponym: "taylorisation" is the breaking-down of some activity into discrete repetitive units, supposedly to improve efficiency. Big Data promises taylorisation on steroids.[108]

[107] I.B. Rich, *A Point for Mr. Taylor*, AM. MACHINIST 674 (Apr. 13, 1911).

[108] Steven Poole, *The Digital Panopticon*, NEW STATESMAN, May 24–30, 2013, at 24, *available at* http://www.newstatesman.com/sci-tech/sci-tech/2013/05/are-you-ready-era-big-data, *archived at* http://perma.cc/2DPN-SBQC; *see also* Steve Johnson, *Hidden Devices Scrutinize Employees*, SAN JOSE MERCURY NEWS (Dec. 13, 2014, 5:58 PM), http://www.mercurynews.com/business/ci_27132016/hidden-devices-scrutinize-employees?source=infinite, *archived at* http://perma.cc/B38Q-KA4B.

[32] Fundamentally, the "predictive" in predictive analytics is to predict human behavior.[109] It can be used, for example, in predicting whether someone has all the skills and resources for a particular job opening.[110] Companies believe they can learn more about a potential employee by scouring the Internet compared to just reading a résumé.[111] Résumés are not considered to be data intensive; instead Internet searches may reveal hints at implicit activity never mentioned in a résumé, such as active participation in relevant online forums or e-mail chat lists.[112] LinkedIn, the popular professional social media site, generates more than half its revenue from selling its data to recruiters.[113] On the flip side, predictive analytics can also be used to decide whether to dismiss an employee.[114]

[33] At present, workplace predictive analytics focus primarily on collecting real-time data to improve productivity under the theory that the more a company knows about its employees the more it can understand their job performance.[115] "You have to bring the same rigor you bring to

[109] *See generally* Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1505 (2013) (discussing government-backed automated predictive initiatives to, for example, predict the spread of pandemics as well as future political and economic developments, based on data collected from Internet traffic, web searches, and Twitter and Facebook posts).

[110] *See* Max Nisen, *Moneyball at Work: They've Discovered What Really Makes A Great Employee*, BUS. INSIDER (May 6, 2013, 1:00 PM), http://www.businessinsider.com/big-data-in-the-workplace-2013-5, *archived at* http://perma.cc/RT8L-BZT7.

[111] *See id.* (referring to resumes as "relics of the dark ages of recruiting").

[112] *See id.* (noting that one startup "looks specifically at the quality of the code engineers put up on GitHub, a popular hosting service for software development projects, to find diamonds in the rough").

[113] *See id.*

[114] *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 162.

[115] *See* Don Reisinger, *Improving Employee Performance With Data Analysis*, CIO INSIGHT (Aug. 20, 2013), http://www.cioinsight.com/it-

operations and finance to the analysis of people."[116] One study of "121 million anonymous performance and behavioral records from . . . a company that provides workforce management information to companies through the use of big data" found the following: only fifty percent of hourly workers will remain with an employer for more than one year; employees who use Chrome or Firefox web browsers instead of Microsoft's Internet Explorer "stay at their jobs longer, miss fifteen percent fewer work days, and deliver higher customer satisfaction;" and "[e]mployees who use three to four social networks are more likely to perform better in their jobs than those who are less involved with social networks."[117] "An employee retention program developed by software company SAS, for example, crunches data on employees who have quit in the past five years—their skills, profiles, studies, and friendships. Then it finds current employees with similar patterns. Another SAS program pinpoints the workers most likely to suffer accidents."[118] Stephen Burks et al. combined personnel data from nine large firms in three industries (call-centers, trucking, and high-tech) that spanned hundreds of thousands of workers and millions of applicants and concluded in part that, while referred applicants have similar skills to non-referred applicants, referred

management/workplace/slideshows/improving-employee-performance-with-data-analysis-09/, *archived at* http://perma.cc/X726-DKEX; *see also* Steve Lohr, *Scientific Management Redux: The Difference Is In The Data*, N.Y. TIMES BLOG (Apr. 21, 2013, 11:29 AM), http://bits.blogs.nytimes.com/2013/04/21/scientific-management-redux-the-difference-is-in-the-data/, *archived at* http://perma.cc/BCD7-NLUA.

[116] Stephen Baker, *Data Mining Moves to Human Resources*, BUSINESSWEEK (Mar. 11, 2009), http://www.bloomberg.com/bw/stories/2009-03-11/data-mining-moves-to-human-resources, *archived at* http://perma.cc/5NRK-CXSW (quoting Rupert Bader, director of workforce planning at Microsoft) (internal quotation marks omitted).

[117] *See* Reisinger, *supra* note 115. Recall that data mining reveals correlations, not necessarily causation. *See supra* text accompanying notes 20–24.

[118] Baker, *supra* note 116.

applicants are most likely to be hired.[119] Marjorie Laura Kane-Sellers applied predictive analytics to fourteen years of sales force retention data of a Fortune 500 company and concluded "training and development participation contributes more significantly to employee retention than salary and job title promotions to the firm's ability to retain sales professionals."[120] And IBM analysts are reportedly charting the skills and experience of IBM's entire workforce in an effort to predict skills that will be needed in the future.[121]

[34] In a slightly different approach, Ben Waber analyzed career outcomes at three U.S. companies using sensor ID badges that monitor physical movement and detect conversations and speech patterns using a combination of infrared, Bluetooth, and microphone data.[122] In addition to the sensor data, the researchers also looked at e-mail, instant message, and phone call data.[123] Waber's results focused on workplace performance versus limited career growth for women: at a banking call center women were measured as more productive than men, but women were disadvantaged when it came to winning promotions and reaching the

[119] *See* STEPHEN BURKS ET AL., THE FACTS ABOUT REFERRALS: TOWARD AN UNDERSTANDING OF EMPLOYEE REFERRAL NETWORKS 2 (2013), *available at* http://ssrn.com/abstract=2253738, *archived at* http://perma.cc/3MLK-YM3Z.

[120] Marjorie Laura Kane-Sellers, Predictive Models of Employee Voluntary Turnover in a North American Professional Sales Force Using Data-Mining Analysis, at iv, 8 (Aug. 2007) (unpublished Ph.D. dissertation, Tex. A&M Univ.), *available at* https://repository.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-1486/KANE-SELLERS-DISSERTATION.pdf, *archived at* https://perma.cc/KMR3-UXYJ.

[121] *See* Baker, *supra* note 116.

[122] *See* Ben Waber, *What Data Analytics Says About Gender Inequality In The Workplace*, BUSINESSWEEK (Jan. 30, 2014), http://www.businessweek.com/articles/2014-01-30/gender-inequality-in-the-workplace-what-data-analytics-says, *archived at* http://perma.cc/9X79-ZNCS (noting the actual content of conversations was ignored to protect employee privacy). *See generally* WABER, *supra* note 10, at 11–12 (discussing wearable sensor badges).

[123] *See* Waber, *supra* note 122.

upper echelons of management; at a pharmaceutical company, women were fractionally more likely to be promoted than men based on the researchers' model, "[y]et only [thirteen] percent of top executives at the company are female despite a 50-50 gender split in the overall workforce."[124]

[35] Meanwhile, at companies embracing workforce analytics, every e-mail, instant message, phone call, line of written code, and mouse-click can be collected and measured.[125] Where a worker eats lunch during the workday may also be monitored and analyzed.[126] Proponents of workplace predictive analytics acknowledge that extensive monitoring "could create a poisonous work environment, one in which people are constantly worried about being spied on and monitored down to the tiniest movement."[127]

[36] Marc Rotenberg, Executive Director of the Electronic Privacy Information Center, suggests the worker is metaphorically behind a one-way mirror: "You don't know what data is being collected and how it is used."[128] And while proponents and practitioners of workplace analytics express concern for worker privacy,[129] as discussed previously,

[124] *Id.*

[125] *Cf.* Steve Lohr, *Big Data, Trying to Build Better Workers*, N.Y. TIMES, Apr. 21, 2013, http://www.nytimes.com/2013/04/21/technology/big-data-trying-to-build-better-workers.html?_r=0, *archived at* http://perma.cc/N4B2-VPN5 (explaining the concept of work-force science and how it is being used in business today).

[126] *See* WABER, *supra* note 10, at 73 (arguing eating lunch with work colleagues is more productive than eating alone at one's desk).

[127] *Id.* at 180.

[128] Lohr, *supra* note 125.

[129] *See, e.g.*, Waber, *supra* note 122 (noting that while examining speech patterns in conversations, researchers "ignored the actual content of conversations to protect privacy").

anonymous data do not necessarily remain anonymous when combined with other data,[130] and non-substantive data (such as phone call numbers dialed, websites visited, e-mail metadata) can reveal just as much personal and private information as the contents of communications themselves.[131]

B. Predictive Analytics and Discrimination

[37] Recall the study mentioned earlier indicating that employees who use Chrome or Firefox web browsers instead of Microsoft's Internet Explorer stay at their jobs longer, miss fifteen percent fewer work days, and deliver higher customer satisfaction.[132] Could this finding prove to be self-fulfilling? In other words, perhaps management begins treating employees using Internet Explorer differently—assuming they are less productive anyway and will probably quit soon—leading to dissatisfaction by those employees, lower productivity, and eventual dismissal. Or perhaps the employer begins using this criterion (use of Internet Explorer) as a hiring factor. However, what if an employee or job applicant uses Internet Explorer because it provides accessibility options that better accommodate the employee's disability? Is the employer discriminating on the basis of disability?

[38] Federal law prohibits discrimination based on race, color, religion, sex, or national origin,[133] disability,[134] age (if forty years or older),[135]

[130] *See supra* text accompanying note 48.

[131] *See* sources cited *supra* note 72.

[132] *See* Reisinger, *supra* note 115.

[133] *See* The Equal Pay Act of 1963, Pub. L. No. 88-38, 77 Stat. 56 (1963) (codified at 29 U.S.C. § 206(d) (2012)); Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241, 253, 255 (1964) (codified 42 U.S.C. §2000e(2) (2012)) [hereinafter "Title VII"]; Pregnancy Discrimination Act of 1978, Pub. L. No. 95-555, 92 Stat. 2076 (1978) (codified at 42 U.S.C. § 2000e(k) (2012)).

[134] *See* Americans with Disabilities Act, Pub. L. No. 101-336, 104 Stat. 327, 337 (1990) (codified at 42 U.S.C.§12112 (2012)).

genetic information,[136] and military service.[137] These laws expressly prohibit employers from making any employment-related decisions based solely or in part on an employee's or job applicant's membership in any of these enumerated "protected classes."[138] However, the Supreme Court

[135] *See* Age Discrimination in Employment Act, Pub. L. No. 90-202, 81 Stat. 602-03 (1967) (codified at 29 U.S.C. §§632, 631(2012)).

[136] *See* Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881, 883 (2008) (codified in scattered sections of 29 U.S.C.).

[137] *See* The Uniformed Services Employment and Reemployment Rights Act of 1994, Pub. L. No. 103-353 § 4301, 108 Stat. 3149, 3150, 3153 (1994) (codified at 38 U.S.C. §4311 (2012)).

[138] *See, e.g.*, 29 U.S.C. § 206(d)(1) (2012) ("No employer . . . shall discriminate . . . between employees on the basis of sex by paying wages to employees . . . at a rate less than the rate at which he pays wages to employees of the opposite sex in such establishment for equal work . . . ."); 29 U.S.C. § 623(a) (2012) ("It shall be unlawful for an employer—(1) to fail or refuse to hire or to discharge any individual or otherwise discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's age . . . ."); 38 U.S.C. § 4311(a) (2012) ("A person who is a member of, applies to be a member of, performs, has performed, applies to perform, or has an obligation to perform service in a uniformed service shall not be denied initial employment, reemployment, retention in employment, promotion, or any benefit of employment by an employer on the basis of that membership, application for membership, performance of service, application for service, or obligation.") 42 U.S.C. § 2000e-2(a) (2012) ("It shall be an unlawful employment practice for an employer—(1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin . . . ."); 42 U.S.C. § 2000e(k) (2012) (For purposes of Title VII, "[t]he terms 'because of sex' or 'on the basis of sex' include, but are not limited to, because of or on the basis of pregnancy, childbirth, or related medical conditions . . . ."); 42 U.S.C. § 2000ff-1(a) (2012) ("It shall be an unlawful employment practice for an employer—(1) to fail or refuse to hire, or to discharge, any employee, or otherwise to discriminate against any employee with respect to the compensation, terms, conditions, or privileges of employment of the employee, because of genetic information with respect to the employee . . . ."); 42 U.S.C. § 12112(a) (2012) ("No covered [employer] shall discriminate against a qualified individual on the basis of disability in regard to job application procedures, the hiring, advancement, or discharge of employees, employee compensation, job training, and other terms, conditions, and privileges of

and Congress have recognized that discrimination can be more subtle—resulting from selection criteria that unintentionally but significantly negatively impacts a protected class. As a result of the Supreme Court's holding that Title VII requires "the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification," [139] Congress amended Title VII to expressly proscribe unintentional (i.e., disparate impact) discrimination. [140] The Supreme Court has also recognized that disparate impact claims apply under the Americans with Disabilities Act [141] and the Age Discrimination in Employment Act.[142]

[39] While arguments have been made that big data is objective, raw data is immune to social bias, and mass-level analysis will avoid group-based discrimination, big data is actually used to segregate individuals into groups.[143] Using big data for employment-related selection criteria can

employment."). All states have parallel laws that prohibit at least Title VII and disability, often with broader definitions of protected classes.

[139] Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971) ("The Act proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation.").

[140] *See* Civil Rights Act of 1991, Pub. L. No. 102-166, § 105(a), 105 Stat. 1071, 1074 (1991) (codified at 42 U.S.C. § 2000e-2(k)(1)(A) (2012)) ("An unlawful employment practice based on disparate impact is established . . . if—(i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity. . . .").

[141] *See* Raytheon Co. v. Hernandez, 540 U.S. 44, 53 (2003) (applying 42 U.S.C. § 12112(b)).

[142] *See* Smith v. City of Jackson, 544 U.S. 228, 240 (2005) (plurality).

[143] *See* Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 10, 2013), http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data, *archived at* http://perma.cc/7GDR-EPZY; *see generally* Peter A. Chow-White & Sandy Green, Jr.,

potentially "reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply reflect the widespread biases that persist in society."[144] Data mining may reflect the quintessential unintentional discrimination, as the adverse selection criteria "may be an artifact of the data mining process itself, rather than a result of programmers assigning certain factors inappropriate weight."[145] Simply not having a LinkedIn profile or the ability to maintain a sophisticated electronic presence may disadvantage certain protected classes.[146]

[40] While predictive analytics have been used to try to root out systemic discrimination,[147] they may just as likely be perpetuating biases. Recall that the goal of some predictive analytics is to select candidates who will remain on the job longer.[148] A recent PriceWaterhouseCoopers

Data Mining Difference in the Age of Big Data: Communication and the Social Shaping of Genome Technologies from 1998 to 2007, 7 INT'L J. COMM. 556, 556 (2013) (asserting that the "seeming neutrality of data mining obfuscates domain assumptions and leaves cultural values and practices of power unexamined").

[144] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV., at *3–4 (forthcoming 2016), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899, *archived at* http://perma.cc/559N-7TUR.

[145] *Id.* at *4.

[146] *See id.* at *14–15 (noting big data analytics may negatively impact "historically disadvantaged groups at higher rates because they are less involved in the formal economy and its data-generating activities, [or] because they have unequal access to and relatively less fluency in the technology necessary to engage online"); Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55, 57 (2013), *available at* http://www.stanfordlawreview.org/sites/default/files/online/topics/66_stanlrevonline_55_lerman.pdf, *archived at* http://perma.cc/R529-45D2 (noting the "the nonrandom, systemic omission of people who live on big data's margins"); *supra* text accompanying notes 111–112.

[147] *See, e.g.*, *supra* text accompanying note 124 (discussing Waber's focus on workplace performance versus limited career growth for women).

[148] *See supra* text accompanying notes 117–118.

survey found that while external hires are increasing, more new hires are leaving within their first year.[149] This is one factor that is driving the increasing use of predictive analytics in hiring.[150] However, if historical data reflect that women leave their job after fewer years compared to men, women may be more likely to be selected out of a predictive optimization analysis.[151]

[41] However, disparate impact discrimination claims are notoriously difficult to establish. Under the disparate-impact statute,[152] a plaintiff establishes a *prima facie* violation by showing that an employer uses "a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin."[153] An employer may defend against liability by demonstrating that the practice is "job related for the position in question and consistent with business necessity."[154] However, even if the employer meets that burden, a plaintiff may still succeed by showing that the employer refuses to adopt an "available alternative employment practice that has less disparate impact and serves the employer's legitimate needs."[155]

---

[149] *See* PRICEWATERHOUSECOOPERS LLP, STATE OF THE WORKFORCE: PWC SARATOGA'S 2013/2014 U.S. HUMAN CAPITAL EFFECTIVENESS REPORT 10 (2013) [hereinafter STATE OF THE WORKFORCE], *available at* http://www.pwc.com/en_US/us/hr-management/publications/assets/pwc-saratoga-human-capital-effectiveness-report.pdf, *archived at* http://perma.cc/N8VB-JTNS; *see also* Andrew R. McIlvaine, *The Power (And Peril) of Predictive Analytics*, HUMAN RES. EXEC. ONLINE (May 21, 2014), http://www.hreonline.com/HRE/view/story.jhtml?id=534357136, *archived at* http://perma.cc/8FQ5-XA4P.

[150] *See* STATE OF THE WORKFORCE, *supra* note 149, at 3.

[151] *See* Barocas & Selbst, *supra* note 144, at *58.

[152] *See supra* note 140.

[153] 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012).

[154] *Id.*

[155] Ricci v. DeStefano, 557 U.S. 557, 578 (2009); *see also* 42 U.S.C. §§ 2000e–2(k)(1)(A)(ii), (k)(1)(C) (2012).

[42] Ironically, to prove a disparate impact case a plaintiff must make a threshold showing of a "significant statistical disparity;" thus "statistics are critical to establish a *prima facie* disparate impact claim . . . ."[156] Although big data and predictive analytics may be more concerned with correlation instead of causation,[157] courts closely scrutinize statistical evidence in disparate impact cases and commonly dismiss them based on statistical error.[158] Disparate impact cases are also relatively expensive, due not only to the need for statistical data and experts,[159] but also because disparate impact is a class-based theory and class action lawsuits have their own significant up-front costs.[160] Finally, disparate impact remedies are fairly limited—they include only substantive remedies, just as job reinstatement or promotion, not compensatory or punitive damages.[161]

[43] Of course, employers will argue that their predictive analytics are directly tied to business needs such as productivity, retention, and promotion. And some may argue their analytics actually correct discrimination by identifying situations in which members of a protected class are not being treated fairly.[162] But with the practical hurdles to successfully pursue a disparate impact case, it is easy to conclude that data mining—and by implication predictive analytics—could potentially

[156] E. Ericka Kelsaw, *Help Wanted: 23.5 Million Unemployed Americans Need Not Apply*, 34 BERKELEY J. EMP. & LAB. L. 1, 22 (2013).

[157] *See supra* text accompanying notes 21–23.

[158] *See* Kelsaw, *supra* note 156, at 23, 28.

[159] *See id.* at 27–28.

[160] *See id.* at 28.

[161] *See id.*

[162] *See, e.g.*, *supra* text accompanying note 124 (discussing Waber's focus on workplace performance versus limited career growth for women).

"exacerbate existing inequalities in difficult-to-counter ways."[163]

C. Limiting Predictive Analytics in the Workplace?

[44] One could argue that screening job applicants or making other job-related decisions based on a variety of outside inputs could fall within the purview of the Fair Credit Reporting Act (FCRA),[164] which very broadly defines a "consumer report" as:

> [A]ny written, oral, or other communication of any information by a consumer reporting agency bearing on a consumer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living which is used or expected to be used or collected in whole or in part for the purpose of serving as a factor in establishing the consumer's eligibility for— . . . (B) employment purposes . . . .[165]

The FCRA defines "employment purposes" as "evaluating a consumer for employment, promotion, reassignment or retention as an employee."[166] The FTC recently investigated the company Social Intelligence, a company that provides pre-employment background screening using information gleaned from the Internet and through social media, concluding the company was in compliance with the FCRA—implying that Social Intelligence's services fell under the auspices of the FCRA.[167]

---

[163] Barocas & Selbst, *supra* note 144, at *59.

[164] *See* Fair Credit Reporting Act, Pub. L. 91-508, Title VI, § 601, 84 Stat. 1114, 1128 (1970) (codified as amended at 15 U.S.C. §§ 1681–1681x (2012)).

[165] 15 U.S.C. § 1681a(d)(1) (2012).

[166] *Id.* § 1681a(h). A "consumer" under the FCRA means an individual. *Id.* § 1681a(c).

[167] *See* Letter from Maneesha Mithal, Associate Director, Federal Trade Commission Division of Privacy and Identity Protection, to Renee Jackson, Att'y for Social Intelligence (May 9, 2011), *available at*

[45] The FCRA does not prevent employers from using background information to vet job applicants and employees;[168] it merely requires certain disclosures[169] and restricts a limited set of information that can be reported.[170] The FCRA also only applies when an employer uses a "consumer reporting agency"[171] to compile the consumer report; as such, it will not apply to larger employers—such as SAS or IBM—that perform their own internal job-related predictive analytics.[172]

[46] The FCRA could provide a foundation for limiting the use of predictive analytics in the workplace. The "mosaic" theory instructs us that using predictive analytics is not necessarily "fair game" just because the underlying data are derived from publicly available information or information in the hands of third parties. A theoretical framework therefore exists to argue that predictive analytics using data mined from

http://www.ftc.gov/sites/default/files/documents/closing_letters/social-intelligence-corporation/110509socialintelligenceletter.pdf, *archived at* http://perma.cc/8HD9-C3EM; *see also* Kashmir Hill, *Social Media Background Check Company Ensures that Job-Threatening Facebook Photos Are Part of Your Application*, FORBES (June 20, 2011, 12:07 PM), http://www.forbes.com/sites/kashmirhill/2011/06/20/now-your-embarrassingjob-threatening-facebook-photos-will-haunt-you-for-seven-years/, *archived at* http://perma.cc/2F75-Q8MG.

[168] However, ten states do limit employers' use of credit background checks. *See Use of Credit Information in Employment 2013 Legislation*, NAT'L CONFERENCE STATE LEGISLATURES, http://www.ncsl.org/research/financial-services-and-commerce/use-of-credit-info-in-employ-2013-legis.aspx, *archived at* http://perma.cc/P5VW-FPZD (last updated Sept. 29, 2014).

[169] *See* 15 U.S.C. §§ 1681d, 1681m(a) (2012).

[170] *See id.* § 1681c (generally restricting information dating back more than seven or ten years).

[171] *Id.* § 1681a(f) (generally defining a consumer reporting agency as a person or entity that regularly compiles consumer reports for third parties for a fee).

[172] *See supra* text accompanying notes 118 & 121.

big data deserve some regulation. At a minimum, notice should be provided to job applicants and employees if adverse decisions are based on analytics.[173] More importantly, since predictive analytics rely more on behavioral and social data—unlike "traditional" consumer reports that are based on discrete fact-based transactions that can be challenged for accuracy—employers should have the burden to establish causation, not just correlation. The mosaic theory can also add a layer of restriction by ensuring that big data profiles created for job applicants or employees do not extend beyond what society and individuals would reasonably expect; that they do not result in an intimate portrait greater than the sum of their parts.[174] In short, every aspect of individuals' digital lives should not be collected and analyzed by employers without restrictions on reasonableness and accuracy (of predictions, not of the underlying data).

[47] Proponents of predictive analytics may argue that such restrictions could stifle progress and potentially prevent employers from using valuable tools that could boost productivity. However, cause and effect must be balanced. Predictive analytics are unproven and subject to error.[175] In the meantime, individuals' livelihoods may be adversely affected. Should we allow individuals to be denied a job, denied a promotion, or even fired, just because of a random correlation based on behavioral information employees had no idea their employers had access to?

V. CONCLUSION

[48] Workforce predictive analytics go beyond mere workplace monitoring. Employers are not just monitoring call length, how many boxes a worker can pack in an hour, or whether a worker is "cyberloafing" by watching cat videos instead of processing invoices. By analyzing

[173] *Cf.* 15 U.S.C. §§ 1681g, 1681m (2012).

[174] *See supra* text accompanying note 72.

[175] *See supra* text accompanying notes 20–22, 24.

social media communications, personal interaction data, even with whom one eats lunch, employers are attempting to build profiles going well beyond the mechanics of work and that can potentially reveal more about a worker than he or she knows about him or herself. The same concerns raised by proponents of the mosaic theory and critiques of data agglomeration apply equally—if not more[176]—to workplace analytics.

[49] The threat to worker privacy is real, so tradeoffs will have to be considered, beginning with data collection standards. Unfortunately, most calls for standards relate to general Internet tracking, are fairly amorphous,[177] and are not workplace-specific. The World Economic Forum has at least made one fairly concrete suggestion, though still related to commercial tracking: govern the usage of data rather than the data themselves.[178] Meanwhile, Ian Kerr and Jessica Earle conclude that "[b]ig data enables a universalizable strategy of preemptive social decision-

[176] *See generally* Lothar Determann & Robert Sprague, *Intrusive Monitoring: Employee Privacy Expectations Are Reasonable in Europe, Destroyed in the United States*, 26 BERKELEY TECH. L.J. 979, 980 (2011) (arguing that U.S. employees have minimal workplace privacy protections principally because employers can defeat any expectation of privacy by notifying employees that intrusive monitoring is taking place). *Cf.* City of Ontario v. Quon, 560 U.S. 746, 760 (2010) ("[E]mployer policies concerning [cell phone and text message] communications will of course shape the reasonable expectations of their employees, especially to the extent that such policies are clearly communicated.").

[177] *See, e.g.*, BIG DATA: SEIZING OPPORTUNITIES, *supra* note 2, at 59 (recommending, in part, maintaining privacy values); WORLD ECON. FORUM, PERSONAL DATA: THE EMERGENCE OF A NEW ASSET CLASS 32–35 (2011), *available at* http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf, *archived at* http://perma.cc/8LL5-9GXU (recommending world leaders should take steps to: "[i]nnovate around user-centricity and trust"; "[d]efine global principles for using and sharing personal data"; "[s]trengthen the dialog between regulators and the private sector"; "[f]ocus on interoperability and open standards"; and "[c]ontinually share knowledge").

[178] *See* World Econ. Forum, *Unlocking the Value of Personal Data: From Collection to Usage* 4 (2013), *available at* http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf, *archived at* http://perma.cc/74VK-9AGZ.

making [that] renders individuals unable to observe, understand, participate in, or respond to information gathered or assumptions made about them;" in other words, "big data can be used to make important decisions that implicate us without our even knowing it."[179] As such, they argue for a reexamination of privacy and due process values—"namely, that there is wisdom in setting boundaries around the kinds of assumptions that can and cannot be made about people."[180] This supports this article's earlier call for disclosure whenever predictive analytics are used in an adverse employment decision and placing the burden on the employer to show causation, not just correlation in the analytics' conclusions.

[50] But perhaps workplace analytics will suffer the same fate as Taylorism before privacy, anti-discrimination law, or other regulations can catch up. In 1910 a group of molders walked off the job at Watertown Arsenal because they refused to be monitored by a supervisor with a stopwatch.[181] The strike led to a five month Congressional hearing to "investigate the Taylor and other systems of shop management."[182] The Committee's report concluded that "[n]either the Taylor system nor any other should be imposed from above on an unwilling work force."[183] "A fair day's work can't be fixed by a stopwatch, which could 'determine the time in which a piece of work can be done, but [not] . . . the length of time in which it ought to be done.'"[184]

[51] Until the courts begin to recognize the threats to privacy by

[179] Kerr & Earle, *supra* note 9, at 71.

[180] *Id.* at 66.

[181] *See generally* KANIGEL, *supra* note 104, at 451–54 (describing the events leading to the molders' strike).

[182] *Id.* at 459.

[183] *Id.* at 482 (internal quotation marks omitted).

[184] *Id.* at 482–83 (alteration in original).

ubiquitous tracking—preferably through a mosaic theory applied not only to private trackers but also particularly to employers—everyone faces the risk of anonymous third parties knowing the intimate details of their private lives. Until then, in the non-workforce environment, we have almost no choice but to succumb to the tracking.[185] Within the workforce, though, today's workers are left with a Hobson's choice of giving up their privacy or giving up their job,[186] if the predictive analytics even allow them to have the job.[187]

[185] *See generally* Goldstein, *supra* note 46 (stating "hiding from big data is inconvenient and expensive"); *supra* text accompanying note 101.

[186] *Cf.* Bauerlein & Jeffery, *supra* note 106, at 20, 23 (noting that modern workers must comply with increased demands at work or risk losing their jobs).

[187] Frank Pasquale urges us "to face the darker possibilities betokened by current trends." FRANK PASQUALE, THE BLACK BOX SOCIETY 16–17 (2015) ("We have come to rely on the titans of reputation, search, and finance to help us make sense of the world; it is time for policymakers to help us make sense of the sensemakers.").