66 STAN. L. REV. ONLINE 65
September 3, 2013

PREDICTION, PREEMPTION, PRESUMPTION: HOW BIG DATA THREATENS BIG PICTURE PRIVACY

Ian Kerr & Jessica Earle*

Big data's big utopia was personified towards the end of 2012.

In perhaps the most underplayed tech moment in the first dozen years of the new millennium, Google brought The Singularity nearer,[1] hiring Ray Kurzweil not as its chief futurist but as its director of engineering. The man the *Wall Street Journal* dubbed a restless genius announced his new post rather quietly in mid-December, without so much as an official press release from Google.[2] This is remarkable when one considers exactly what Google hired him to do. Kurzweil and his team will try to create a mind—an artificial intellect capable of predicting on a "semantically deep level what you are interested in."[3] With easy access to the search giant's enormous user base and the potential to scour all Google-mediated content, Kurzweil (and apparently Google) aims to turn the very meaning of "search" on its head: instead of people using

[*] Ian Kerr is Canada Research Chair in Ethics, Law and Technology, Faculty of Law, University of Ottawa. Jessica Earle is JD/MA Candidate, Faculty of Law, University of Ottawa and Norman Paterson School of International Affairs, Carleton University.

[1] Or, not. *See* John Pavlus, *By Hiring Kurtzweil, Google Just Killed the Singularity*, MIT TECH. REV. (Dec. 17, 2012), http://www.technologyreview.com/view/508901/by-hiring-kurzweil-google-just-killed-the-singularity.

[2] William M. Bulkeley, *Kurzweil Applied Intelligence Inc.*, WALL ST. J., June 23, 1989, at A3. In an email sent on July 7, 2013, Google confirmed that the hire was not announced by the search giant, but was posted on Kurzweil's website, http://www.kurzweilai.net/kurzweil-joins-google-to-work-on-new-projects-involving-machine-learning-and-language-processing. E-mail from Jason Freidenfelds, Google Communications Representative, to Jessica Earle (July 7, 2013, 17:52 UTC) (on file with Stanford Law Review).

[3] Interview by Keith Kleiner with Ray Kurzweil, Director of Engineering, Google, in Moffett Field, Cal. (Jan. 4, 2013), *available at* http://www.youtube.com/watch?v=YABUffpQY9w.

search engines to better understand information, search engines will use big data to better understand people. As Kurzweil has characterized it, intelligent search will provide information to users before they even know they desire it. This accords precisely with Larry Page's longstanding vision: intelligent search "understands exactly what you mean and gives you back exactly what you want."[4]

Kurzweil's new project reifies society's increasing optimism in harnessing the utility of big data's predictive algorithms—the formulaic use of zetabytes of data to anticipate everything from consumer preferences and customer creditworthiness to fraud detection, health risks, and crime prevention. Through the predictive power of these algorithms, big data promises opportunities like never before to anticipate future needs and concerns, plan strategically, avoid loss, and manage risk. Big data's predictive tool kit clearly offers many important social benefits.[5] At the same time, its underlying ideology also threatens fundamental legal tenets such as privacy and due process.

Contrary to the received view, our central concern about big data is *not* about the data. It is about big data's power to enable a dangerous new philosophy of preemption. In this Essay, we focus on the social impact of what we call "preemptive predictions." Our concern is that big data's promise of increased efficiency, reliability, utility, profit, and pleasure might be seen as the justification for a fundamental jurisprudential shift from our current ex post facto system of penalties and punishments to ex ante preventative measures that are increasingly being adopted across various sectors of society. It is our contention that big data's predictive benefits belie an important insight historically represented in the presumption of innocence and associated 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.[6]

I. PREDICTION

Since much of the big data utopia is premised on prediction, it is important to understand the different purposes that big data predictions serve. This Part offers a quick typology.

The nature of all prediction is anticipatory. To predict is to "state or estimate . . . that an action or event will happen in the future or will be a conse-

[4] *Our Products and Services*, GOOGLE, http://www.google.com/corporate/tech.html (last visited Aug. 29, 2013) (internal quotation marks omitted).

[5] Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 Stan. L. Rev. Online 63, 64 (2012).

[6] Our argument in this brief Essay is an adaptation of an earlier book chapter, Ian Kerr, *Prediction, Pre-emption, Presumption: The Path of Law After the Computational Turn*, *in* PRIVACY, DUE PROCESS AND THE COMPUTATIONAL TURN 91 (Mireille Hildebrandt & Katja de Vries eds., 2013).

quence of something."[7] For example, when a lawyer predicts "what the courts will do in fact,"[8] she anticipates the legal consequences of future courses of conduct in order to advise clients whether it is feasible to avoid the risk of state sanction. We call predictions that attempt to anticipate the likely consequences of a person's action *consequential predictions.* As doctors, lawyers, accountants, and other professional advisors are well aware, the ability to make reliable consequential predictions can be profitable—especially in a society increasingly preoccupied with risk. The recent development of anticipatory algorithms within these fields is generally client centered. [9] The aim of these prediction services is to allow individuals to eschew risk by choosing future courses of action that best align with their own self-interest, forestalling unfavorable outcomes.

Of course, not all of big data's predictions are quite so lofty. When you permit iTunes Genius to anticipate which songs you will like or Amazon's recommendation system to predict what books you will find interesting, these systems are not generating predictions about your conduct or its likely consequences. Rather, they are trying to stroke your preferences in order to sell goods and services. Many of today's big data industries are focused on projections of this material sort, which we refer to as *preferential predictions.* Google's bid to create personalized search engines is a prime example of society's increasing reliance on preferential predictions. The company's current interface already uses anticipatory algorithms to predict what information users want based on a combination of data like website popularity, location, and prior search history.

There is a third form of prediction exemplified by a number of emerging players in big data markets. Unlike consequential and preferential predictions, *preemptive predictions* are intentionally used to diminish a person's range of future options. Preemptive predictions assess the likely consequences of allowing or disallowing a person to act in a certain way. In contrast to consequential or preferential predictions, preemptive predictions do not usually adopt the perspective of the actor. Preemptive predictions are mostly made from the standpoint of the state, a corporation, or anyone who wishes to prevent or forestall certain types of action. Preemptive predictions are not concerned with an individual's actions but with whether an individual or group should be permitted to act in a certain way. Examples of this technique include a no-fly list used to preclude possible terrorist activity on an airplane, or analytics software used to determine how much supervision parolees should have based on predic-

[7] *See Predict Definition*, OXFORD ENGLISH DICTIONARY, http://www.oed.com/view/Entry/149856 (last visited Aug. 29, 2013).

[8] Oliver W. Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 461 (1897).

[9] *See IBM Watson*, IBM, http://www-03.ibm.com/innovation/us/watson (last visited Aug. 29, 2013); *see also AI Am the Law*, ECONOMIST (Mar. 10, 2005), http://www.economist.com/search/PrinterFriendly.cfm?story_id=3714082.

tions of future behavior.[10] The private sector is also embracing this approach. For example, companies are increasingly combing through big data to find their job candidates, rather than looking to the traditional format of resumes and interviews.[11]

These three types of prediction—consequential, preferential, and preemptive—are not meant to provide an exhaustive list of all possible predictive purposes. But, as the following section reveals, understanding the different predictive purposes will help locate the potential threats of big data. To date, much of the academic focus on big data and privacy investigates what we have called consequential and preferential predictions in the context of data protection frameworks.[12] In this Essay, we focus on the less understood category of preemptive prediction and its potential impact on privacy and due process values.

II. PREEMPTION

The power of big data's preemptive predictions and its potential for harm must be carefully understood alongside the concept of risk. When sociologist Ulrich Beck coined the term *risk society* in the 1990s, he was not suggesting that society is riskier or more dangerous nowadays than before; rather, he argued that society is reorganizing itself in response to risk. Beck believes that in modern society, "the social production of wealth is systematically accompanied by the social production of risks," and that, accordingly, "the problems and conflicts relating to distribution in a society of scarcity overlap with the problems and conflicts that arise from the production, definition, and distribution of techno-scientifically produced risks."[13]

On Beck's account, prediction and risk are interrelated concepts. He subsequently describes risk as "the modern approach to foresee and control the future consequences of human action . . . ."[14] This helps to demonstrate the link

[10] Soumya Panda, *The Procedural Due Process Requirements for No-Fly Lists*, 4 PIERCE L. REV. 121 (2005); Steve Watson, *Pre-Crime Technology to Be Used in Washington D.C.*, PRISON PLANET (Aug. 24, 2010), http://www.prisonplanet.com/pre-crime-technology-to-be-used-in-washington-d-c.html.

[11] *E.g.*, 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.

[12] *E.g.*, Asim Ansari et al., *Internet Recommendation Systems*, 37 J. MARKETING RES. 363 (2000); Tam Harbert, *Big Data Meets Big Law: Will Algorithms Be Able to Predict Trial Outcomes?*, LAW TECH. NEWS (Dec. 27, 2012), http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202555605051; Ernan Roman, *BIG Data Must Create BIG Experiences*, DIRECT MKTG. NEWS (Mar. 18, 2013), http://www.dmnews.com/big-data-must-create-big-experiences/article/284831; Daniel Martin Katz, Remarks at Michigan State University College of Law's lawTechCamp: Quantitative Legal Prediction (Or How I Learned to Stop Worrying and Embrace Disruptive Technology) (June 8, 2013), *available at* http://lawtechcamp.com/qualitative-legal-prediction.

[13] ULRICH BECK, RISK SOCIETY: TOWARDS A NEW MODERNITY 19 (1992).

[14] ULRICH BECK, WORLD RISK SOCIETY 3 (1999).

between prediction and preemption. Prediction industries flourish in a society where anyone and anything can be perceived as a potential threat, because it is lucrative to exploit risk that can later be avoided. In such cases, prediction often precipitates the attempt to preempt risk.

With this insight, an important concern arises. Big data's escalating interest in and successful use of preemptive predictions as a means of avoiding risk becomes a catalyst for various new forms of social preemption. More and more, governments, corporations, and individuals will use big data to preempt or forestall activities perceived to generate social risk. Often, this will be done with little or no transparency or accountability. Some loan companies, for example, are beginning to use algorithms to determine interest rates for clients with little to no credit history, and to decide who is at high risk for default. Thousands of indicators are analyzed, ranging from the presence of financially secure friends on Facebook to time spent on websites and apps installed on various data devices. Governments, in the meantime, are using this technique in a variety of fields in order to determine the distribution of scarce resources such as social workers for at-risk youth or entitlement to Medicaid, food stamps, and welfare compensation.[15]

Of course, the preemption strategy comes at a significant social cost. As an illustration, consider the practice of using predictive algorithms to generate no-fly lists. Before the development of many such lists in various countries, high-risk individuals were generally at liberty to travel—unless the government had a sufficient reason to believe that such individuals were in the process of committing an offense. In addition to curtailing liberty, a no-fly list that employs predictive algorithms preempts the need for any evidence or constitutional safeguards. Prediction simply replaces the need for proof.

Taken to its logical extreme, the preemption philosophy is not merely proactive—it is aggressive. As President George W. Bush famously argued:

> If we wait for threats to fully materialize, we will have waited too long. . . . We must take the battle to the enemy, disrupt his plans, and confront the worst threats before they emerge. . . . [O]ur security will require all Americans to be forward-looking and resolute, to be ready for preemptive action when necessary . . . .[16]

Proponents of this approach argue there is a "duty to prevent," which means the responsible choice requires use of predictive tools to mitigate future

[15] Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1256 (2008); Stephen Goldsmith, *Big Data, Analytics and a New Era of Efficiency in Government*, GOVERNING THE STATES AND LOCALITIES (May 22, 2013), http://www.governing.com/blogs/bfc/col-big-data-analytics-government-efficiency.html; Evgeny Morozov, *Your Social Networking Credit Score*, Slate (Jan. 30, 2013, 8:30 AM), http://www.slate.com/articles/technology/future_tense/2013/01/wonga_lenddo_lendup_big_data_and_social_networking_banking.html.

[16] President George W. Bush, Graduation Speech at West Point (June 1, 2002, 9:13 AM), *available at* http://georgewbush-whitehouse.archives.gov/news/releases/2002/06/20020601-3.html.

risk. [17] But with this, we see that a universalized preemption strategy could challenge some of our most fundamental jurisprudential commitments, including the presumption of innocence. In the following Part, we seek to demonstrate that even more mundane forms of preemption generated by big data can also threaten privacy and due process values.

III. PRESUMPTION

To date, much of the best work on the implications of big data tends to treat the privacy worry as though it were somehow contained within the minutiae of the data itself. As Tene and Polonetsky have meticulously argued: "Information regarding individuals' health, location, electricity use, and online activity is exposed to scrutiny, raising concerns about profiling, discrimination, exclusion, and loss of control."[18] Through the fine-tuned microscope of data privacy frameworks, the central issues tend to be the definition of personally identifiable information, the prospect of de-identifying the data, the nature of consent to the collection, use, or disclosure of the data, and a range of other data privacy rules such as purpose limitation and data minimization.

Our approach examines the privacy issue with a telescope rather than a microscope.

If the legal universe has a prime directive, it is probably the shared understanding that everyone is presumed innocent until proven guilty. In legal discourse, the presumption of innocence is usually construed, narrowly, as a procedural safeguard enshrined within a bundle of "due process" rights in criminal and constitutional law. These include the right to a fair and impartial hearing, an ability to question those seeking to make a case against you; access to legal counsel, a public record of the proceedings, published reasons for the decision, and, in some cases, an ability to appeal the decision or seek judicial review.[19] Likewise, a corollary set of duties exists in the private sector. Although such duties are not constitutionally enshrined, companies do owe employees and customers the right to full information, the right to be heard, the right to ask questions and receive answers, and the right of redress.[20] Gazing at the bigger picture, the presumption of innocence and related private sector due process values can be seen as wider moral claims that overlap and interrelate with core privacy values.

Taken together, privacy and due process values seek to limit what the government (and, to some extent, the private sector) is permitted to presume about

[17] Lee Feinstein & Anne-Marie Slaughter, *Duty to Prevent*, FOREIGN AFF., Jan.-Feb. 2004, at 136.

[18] Tene & Polonetsky, *supra* note 5, at 65.

[19] Henry J. Friendly, *"Some Kind of Hearing"*, 123 U. PA. L. REV. 1267 (1975).

[20] Kerr, *supra* note 6, at 108. *See generally* Lauren B. Edelman, *Legal Environments and Organizational Governance: The Expansion of Due Process in the American Workplace*, 95 AM. J. SOC., 1401, 1405-08 (1990).

individuals absent evidence that is tested in the individuals' presence, with their participation. As such, these values aim to provide fair and equal treatment to all by setting boundaries around the kinds of assumptions that can and cannot be made about people. This is wholly consistent with privacy's general commitment to regulating what other people, governments, and organizations are permitted to know about us. Among other things, the aim is to prevent certain forms of unwarranted social exclusion.[21]

With all of this, we are finally able to locate the threat that big data poses. Big data enables a universalizable strategy of preemptive social decision-making. Such a strategy renders individuals unable to observe, understand, participate in, or respond to information gathered or assumptions made about them. When one considers that big data can be used to make important decisions that implicate us without our even knowing it, preemptive social decision making is antithetical to privacy and due process values.

CONCLUSION

The nexus between big data and privacy is not a simple story about how to tweak existing data protection regimes in order to "make ends meet"; big data raises a number of foundational issues. Since predictability is itself an essential element of any just decisionmaking process, our contention is that it must be possible for the subjects of preemptive predictions to scrutinize and contest projections and other categorical assumptions at play within the decision-making processes themselves. This is part of our broader assertion that privacy and due process values require setting boundaries around the kinds of institutional assumptions that can and cannot be made about people, particularly when important life chances and opportunities hang in the balance.

We believe that such considerations will become increasingly significant in both public and private sector settings, especially in light of the kinds of big data prediction machines that Ray Kurzweil and others want to build "to . . . Google scale."[22] These projects must be kept in mind given our emerging understanding that "some uses of probability and statistics serve to reproduce and reinforce disparities in the quality of life that different sorts of people can hope to enjoy."[23]

While it is exciting to think about the power of big data and the utopic allure of powerful prediction machines that understand exactly what we mean and tell us exactly what we want to know about ourselves and others, we believe that privacy values merit the further study and development of potential

[21] OSCAR H. GANDY JR., THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION (1993); Richard V. Ericson, *The Decline of Innocence*, 28 U.B.C. L. REV. 367 (1994).

[22] Interview with Ray Kurzweil by Keith Kleiner, *supra* note 3.

[23] OSCAR H. GANDY, JR., COMING TO TERMS WITH CHANCE: ENGAGING RATIONAL DISCRIMINATION AND CUMULATIVE DISADVANTAGE 1 (2009).

limitations on how big data is used. We need to ensure that the convenience of useful prediction does not come at too high a cost.