Prediction, Preemption, Presumption:
How Big Data Threatens Big Picture Privacy

Ian Kerr[*] and 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 "the restless genius" announced his new post rather quietly in mid-December, without so much as an official press release from Google. 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."[2] Driven by big data and the eager permission of its enormous user-base to scour all Google-mediated content, action and interaction, Kurzweil (and apparently Google) aims to turn the very meaning of "search" on its head: instead of people using 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 want it. Intelligent search "understands exactly what you mean and gives you back exactly what you want."[3]

Kurzweil's new project reifies society's increasing willingness to embrace big data's predictive algorithms—the use of formulas to anticipate everything from consumer preferences, customer retention and creditworthiness, through to fraud detection, health risks, crime prevention and genetic inheritance. While big data's predictive tool kit promises many social benefits,[4] we argue that its underlying ideology threatens big picture legal concepts such as privacy and due process by enabling a dangerous new philosophy of preemption.

Contrary to the received view, our concern about big data is *not* about the data.

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* systems of penalties and punishments to *ex ante* preventative measures that are increasingly being adopted across various sectors of society. Efficiency and the bottom line are laudable social goals. However, it is our contention that big data's predictive utopia ignores an important insight historically represented in the presumption of innocence and associated privacy and due process

[*] Canada Research Chair in Ethics, Law and Technology, University of Ottawa, Faculty of Law
[**] JD/MA Candidate, University of Ottawa, Faculty of Law/ Carleton University, Norman Paterson School of International Affairs
[1] Or, not: http://www.technologyreview.com/view/508901/by-hiring-kurzweil-google-just-killed-the-singularity/.
[2] Interview with Ray Kurzweil, Director of Engineering, Google, in Moffett Field, Cal. (Jan 10, 2013), *at* http://www.youtube.com/watch?v=YABUffpQY9w.
[3] *What we do for you*, GOOGLE, http://www.google.com/corporate/tech.html (last visited July 3, 2013).
[4] Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 64 (2012).

values—namely, that there is wisdom in setting boundaries around the kinds of assumptions that can and cannot be made about people.[5]

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 section offers a quick typology.

The nature of all prediction is anticipatory. To predict is to say or assume something before it happens.[6] For example, in the classic Holmesian sense, when a lawyer predicts "what the courts will do in fact,"[7] she anticipates the legal consequences of future courses of conduct in order to advise clients whether it is feasible or desirable 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 lucrative—especially in a society increasingly preoccupied with risk. The recent development of anticipatory algorithms within these fields[8] is generally client-centered. 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 that are not to their advantage.

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 you stuff. 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 used to intentionally diminish a person's range of future options. Preemptive predictions assess the likely consequences of (dis)allowing a person to act in a certain way. Unlike 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 whether an individual or group should be permitted to act in a certain way. Examples of preemptive prediction techniques include a no-fly list used to preclude possible

---

[5] Our argument in this brief article is an adaptation of an earlier book chapter: Ian Kerr, *Prediction, Presumption, Preemption: The Path of Law After the Computational Turn in* PRIVACY, DUE PROCESS AND THE COMPUTATIONAL TURN: THE PHILOSOPHY OF LAW MEETS THE PHILOSOPHY OF TECHNOLOGY 91 (Mireille Hildebrandt & Ekaterina De Vries, eds., 2013).

[6] THE OXFORD ENGLISH DICTIONARY (2d ed.).

[7] Oliver Wendell Holmes, Jr., *The Path of the Law*, *reprinted in* THE COLLECTED WORKS OF JUSTICE HOLMES 291 (Sheldon M. Novick ed., 1995).

[8] *See IBM Watson: Ushering in a New Era of Computing*, http://www-03.ibm.com/innovation/us/watson/; *See also AI am the law*, ECONOMIST, Mar. 10, 2005, http://www.economist.com/search/PrinterFriendly.cfm?story_id=3714082.

terrorist activity on an airplane, or analytics software used to determine how much supervision parolees should have based on predictions of future behavior.[9] The private sector is also embracing this approach, with companies increasingly combing through big data to find their job candidates, rather than looking to the traditional format of resumes and interviews.

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 literature on big data and privacy investigates what we have called consequential and preferential predictions in the context of data protection frameworks.[10] In this article, we focus on the less understood category of preemptive prediction and its potential impact on big picture privacy and due process values.

PREEMPTION

The power of big data's preemptive predictions and their 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 more risky or dangerous nowadays than before; rather, he argued that society is re-organizing 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.[11]

On Beck's account, prediction and risk are interrelated concepts. He subsequently defines risk as "the modern approach to foresee and control the future consequences of human action."[12]

Put bluntly, prediction industries flourish in a society where anyone and anything can be perceived as a potential threat. Here, prediction and preemption go hand in hand. This is because prediction often precipitates the attempt to preempt risk.

---

[9] Richard Berk, UNIVERSITY OF PENNSYLVANIA (Jan. 3, 2011), http://www-stat.wharton.upenn.edu/~berkr/; 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; Max Nisen, *Moneyball at Work: They've Discovered What Really Makes A Great Employee*, BUSINESS INSIDER, May 6, 2013, http://www.businessinsider.com/big-data-in-the-workplace-2013-5#ixzz2YNZGfSXW.

[10] *E.g.*, Asim Ansari, Skander Essegaier & Rajeev Kohli, *Internet Recommendation Systems*, 37 No. 3 J. OF MARKETING RESEARCH 363 (2000); Ernan Roman, *Big Data Must Create BIG Experiences*, DIRECT MARKETING NEWS, March 18, 2013, http://www.dmnews.com/big-data-must-create-big-experiences/article/284831/; Tam Harbert, *Big Data Meets Big Law: Will Algorithms be able to Predict Trial Outcomes?*, LAW TECHNOLOGY NEWS, December 27, 2012, http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202555605051; Daniel Martin Katz, *Quantitative Legal Prediction (Or How I learned to Stop Worrying and Embrace Disruptive Technology* (lecture taught at Michigan State University College of Law Tech Camp on June 8, 2013), http://lawtechcamp.com/qualitative-legal-prediction/.

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

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

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.[13]

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 such lists, high-risk individuals were still at liberty to travel—unless the government had probable cause to believe that such individuals were in the process of committing an offence. In addition to curtailing liberty, a no-fly list 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...our security will require all Americans to be forward-looking and resolute, to be ready for preemptive action when necessary…[14]

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 section, we seek to demonstrate that even more mundane forms of preemption generated by big data can also threaten big picture privacy and due process values.

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 minutia of the data itself. As Tene and Polonetsky have meticulously argued: "[i]nformation regarding individuals' health, location, electricity use, and online activity is exposed to scrutiny, raising concerns about profiling, discrimination, exclusion, and loss of control."[15] Through the fine-tuned microscope of data privacy frameworks,

---

[13] Evgeny Morozov**,** *Your Social Networking Credit Score: "Big data" can help determine who really deserves a loan, but there are dangers,* SLATE, Jan. 30, 2013, http://www.slate.com/articles/technology/future_tense/2013/01/wonga_lenddo_lendup_big_data_and_social_networking_banking.html; *Big Data, Analytics and a New Era of Efficiency in Government,* GOVERNING THE STATE AND LOCALITIES, May 22, 2013, http://www.governing.com/blogs/bfc/col-big-data-analytics-government-efficiency.html; D.K. Citron, *Technological Due Process,* 85 WASH. U. L. REV.1249, 1256 (2007-2008).

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

[15] *Supra* note 4 at 65.

the central issues tend to focus on: 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.[16] Likewise, a corollary set of duties exists in the private sector, with companies under a duty to fulfill employees and customers: right to full information; right to be heard; right to ask questions and receive answers; and right of redress.[17]

Gazing at the bigger picture, the presumption of innocence and related 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 individuals absent evidence that is tested in their 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 individuals. 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.[18]

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 that can circumvent the legal and moral requirement to permit individuals the ability to observe, understand, participate in and respond to important information gathered about them or used in order to make decisions or partake in actions that implicate them. Operationalized in this manner, 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 decision-making 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

[16] H.J. Friendly, *Some Kind of Hearing*, 123 U. PA. L. REV., 1267-1317 (1974-1975).
[17] *Id.*
[18] OSCAR H. GANDY JR., THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION (1993); R.V. Ericson, *The Decline of Innocence*, 28 U. BRIT. COL. L. REV., 367 (1994).

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 the Google scale."[19] Keeping these projects in mind is especially important given our emerging understanding that, "[t]he application of probability and statistics to an ever-widening number of life-decisions serves to reproduce, reinforce, and widen disparities in the quality of life that different groups of people can enjoy."[20]

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 big picture privacy values merit the further study and development of potential 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.

[19] *Supra* note 2 at 4:10.
[20] OSCAR H. GANDY JR., COMING TO TERMS WITH CHANCE: ENGAGING RATIONAL DISCRIMINATION AND CUMULATIVE DISADVANTAGE 242 (2009).