Boston College Law Review

Volume 55 | Issue 1 Article 4

1-29-2014

Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms

Kate Crawford
Microsoft Research, kate@katecrawford.net

Jason Schultz
NYU School of Law, jason.schultz@exchange.law.nyu.edu

BIG DATA AND DUE PROCESS: TOWARD A FRAMEWORK TO REDRESS PREDICTIVE PRIVACY HARMS

KATE CRAWFORD*
JASON SCHULTZ**

Abstract: The rise of "Big Data" analytics in the private sector poses new challenges for privacy advocates. Through its reliance on existing data and predictive analysis to create detailed individual profiles, Big Data has exploded the scope of personally identifiable information ("PII"). It has also effectively marginalized regulatory schema by evading current privacy protections with its novel methodology. Furthermore, poor execution of Big Data methodology may create additional harms by rendering inaccurate profiles that nonetheless impact an individual's life and livelihood. To respond to Big Data's evolving practices, this Article examines several existing privacy regimes and explains why these approaches inadequately address current Big Data challenges. This Article then proposes a new approach to mitigating predictive privacy harms—that of a right to procedural data due process. Although current privacy regimes offer limited nominal due process-like mechanisms, a more rigorous framework is needed to address their shortcomings. By examining due process's role in the Anglo-American legal system and building on previous scholarship about due process for public administrative computer systems, this Article argues that individuals affected by Big Data should have similar rights to those in the legal system with respect to how their personal data is used in such adjudications. Using these principles, this Article analogizes a system of regulation that would provide such rights against private Big Data actors.

INTRODUCTION

Big Data analytics have been widely publicized in recent years, with many in the business and science worlds focusing on how large datasets can

* Principal Researcher, Microsoft Research; Visiting Professor, MIT Centre for Civic Media; Senior Fellow, NYU Information Law Institute.

** Associate Professor of Clinical Law, NYU School of Law. The authors wish to thank Danielle Citron, Michael Froomkin, Brian Pascal, Brian Covington, and the participants in the 2013 Privacy Law Scholars Conference for their valuable feedback. They also wish to thank Stephen Rushin, Ph.D. for his invaluable research assistance.

offer new insights into previously intractable problems.[1] At the same time, Big Data poses new challenges for privacy advocates. Unlike previous computational models that exploited known sources of personally identifiable information ("PII") directly, such as behavioral targeting,[2] Big Data has radically expanded the range of data that can be personally identifying.[3] By primarily analyzing metadata, such as a set of predictive and aggregated findings, or by combining previously discrete data sets, Big Data approaches are not only able to manufacture novel PII, but often do so outside the purview of current privacy protections.[4] Existing regulatory schema appear incapable of keeping pace with these advancing business norms and practices.

Personal harms emerge from the inappropriate inclusion and predictive analysis of an individual's personal data without their knowledge or express consent. For example, in 2012, a well-publicized *New York Times* article revealed that the retail chain Target had used data mining techniques to predict which female customers were pregnant, even if they had not yet announced it publicly.[5] This activity resulted in the unauthorized disclosure of personal information to marketers.[6] In essence, Target's predictive analytics

---

[1] *See, e.g.*, David Brooks, Op-Ed., *The Philosophy of Data*, N.Y. TIMES, Feb. 5, 2013, at A23 (highlighting Big Data's potential); *Strata 2013 Is a Wrap*, STRATA CONF.: MAKING DATA WORK, http://strataconf.com/strata2013, *archived at* http://perma.cc/8KYZ-FPGQ (last visited Dec. 2, 2013) (documenting an internationally recognized Big Data conference). Organizations may use Big Data to combine and analyze large datasets to discover correlations and make predictions. For example, the United Postal Service collects vehicle data from sensors on each of its many delivery trucks and uses predictive algorithms to delegate preventive maintenance—resulting in lower maintenance costs and ensuring timely delivery. Greg Satell, *Yes, Big Data Can Solve Real World Problems*, FORBES (Dec. 3, 2013, 1:45 AM), http://www.forbes.com/sites/gregsatell/2013/12/03/yes-big-data-can-solve-real-world-problems/, *archived at* http://perma.cc/VSW4-QVG3. In addition, Big Data has had a positive impact on the health care industry, with the shift toward "informational research that analyzes large data and biological sample sets" leading to "significant discoveries, the development of new therapies, and a remarkable improvement in health care and public health." *See* Paul M. Schwartz, *Information Privacy in the Cloud*, 161 U. PA. L. REV. 1623, 1631 (2013) (quoting INST. OF MED. OF THE NAT'L ACADS., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 113 (Sharyl J. Nass et al. eds., 2009)).

[2] *See* Elspeth A. Brotherton, *Big Brother Gets a Makeover: Behavioral Targeting and the Third-Party Doctrine*, 61 EMORY L.J. 555, 558 (2012) (describing "behavioral targeting" as "an online advertising technique designed to deliver specific, targeted advertisements to Internet users based on their perceived interests," and observing that companies are able to do this "by using sophisticated technology that tracks and gathers information about users' online activity).

[3] *See* Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 75–76 (2013).

[4] *See id.* at 76–77, 82–83; Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 65–66 (2012), http://www.stanfordlawreview.org/online/privacy-paradox/big-data, *archived at* http://perma.cc/U6ZQ-PSK6.

[5] Charles Duhigg, *Psst, You in Aisle 5*, N.Y. TIMES, Feb. 19, 2012, § 6 (Magazine), at 30.

[6] *Id*.

"guessed" that a customer was pregnant and disclosed her name to their marketing department, manufacturing PII about her instead of collecting it directly.[7] Although the customers likely knew that Target collected data on their individual purchases, it is doubtful that many considered the risk that Target would use data analytics to create such personal customer models to send advertising material to homes. These types of harms do not necessarily fall within the conventional invasion of privacy boundaries, but such harms are still derived from collecting and using information that centers on an individual's data behaviors. We call these "predictive privacy harms."

This Article confronts the tension between the powerful potential benefits of Big Data and the resulting predictive privacy harms. Part I discusses the nature of "Big Data science" and how personal information can be amassed and analyzed.[8] It then discusses the nature of predictive privacy harms and why traditional privacy protections are insufficient to address the risks posed by Big Data's use of personal information.[9] In Part II, this Article recounts the Anglo-American history of procedural due process and the role it has played in both systems of adjudication and separation of powers.[10] Part II then makes the case for why procedural due process principles may be an appropriate source to draw from to address the risks of predictive privacy harms.[11] Finally, Part III looks at the procedural due process literature and suggests ways to analogize a similar framework for private sector Big Data systems.[12]

I. PREDICTIVE PRIVACY HARMS AND THE MARGINALIZATION OF TRADITIONAL PRIVACY PROTECTIONS

A. *What Is Big Data and Why All the Hype?*

> *Knowledge is invariably a matter of degree: you cannot put your finger upon even the simplest datum and say "this we know". In the growth and construction of the world we live in, there is no one stage, and no one aspect, which you can take as the foundation.*
>
> —T.S. Eliot[13]

---

[7] *Id.*

[8] *See infra* notes 14–19 and accompanying text.

[9] *See infra* notes 20–88 and accompanying text.

[10] *See infra* notes 90–172 and accompanying text.

[11] *See infra* notes 126–129, 145–147, 153–172 and accompanying text.

[12] *See infra* notes 173–201 and accompanying text.

[13] T.S. ELIOT, KNOWLEDGE AND EXPERIENCE IN THE PHILOSOPHY OF F.H. BRADLEY 151 (1964).

"Big Data" is a generalized, imprecise term that refers to the use of large data sets in data science and predictive analytics.[14] In practice, the term encompasses three aspects of data magnification and manipulation.[15] First, it refers to technology that maximizes computational power and algorithmic accuracy.[16] Second, it describes types of analyses that draw on a range of tools to clean and compare data.[17] Third, it promotes the belief that large data sets generate results with greater truth, objectivity, and accuracy.[18] The promise of Big Data's ability to analyze data and provide novel insights has led to profound investment in, consideration of, and excitement about Big Data's power to solve problems in numerous disciplines and business arenas.[19]

B. *Big Data's Predictive Privacy Harms*

Alongside its great promise, Big Data presents serious privacy problems. It gathers its contents from a myriad of online user interactions and infrastructure sensors, ranging from online transactions, search queries, and health records to communication networks, electric grids, and mobile phones.[20] Not only are the generated data sets sizable, but they also often contain very intimate aspects of individuals' lives.[21] This Section begins by discussing the expanded scope and quantity of personal information vulnerable to Big Data and concludes by illustrating Big Data's potential harms of enabling discriminatory housing practices, exposing sensitive health information, and facilitating predictive policing.[22]

---

[14] *See* 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) (defining Big Data to include personal data generated from a variety of sources).

[15] *See* danah boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO. COMM. & SOC'Y 662, 663 (2012). Big Data raises numerous critical questions about all three of these uses. *Id.*

[16] *Id.*

[17] *Id.*

[18] *Id.*

[19] *See* Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 388 (2012) (citing JAMES MANYIKA ET AL., MCKINSEY GLOBAL INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY 1 (2011), http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, *archived at* http://perma.cc/7XAZ-QF9J).

[20] Tene & Polonetsky, *supra* note 14, at 240 (noting other sources of Big Data such as email, video, clickstream, logs, social networking interactions, global positioning satellites, roads, bridges, homes, and clothing).

[21] *See* Jay Stanley, *Eight Problems with "Big Data,"* FREE FUTURE, (Apr. 25, 2012, 3:06 PM), http://www.aclu.org/blog/technology-and-liberty/eight-problems-big-data, *archived at* http://perma.cc/RF4U-VF8A.

[22] *See infra* notes 23–40 and accompanying text (expanded scope); *infra* notes 41–74 and accompanying text (potential harms).

The many sources for retrieving and generating data have expanded the amount and availability of personal data.[23] For example, health data is particularly vulnerable; a single breach risks exposing critical information from a multitude of patients' records.[24] Furthermore, as one health information law scholar observes, data about our online behavior generally—such as buying an e-book about breast cancer survival or "liking" a disease foundation's Facebook page—can also reveal information about our health.[25] Even the radio-frequency identification (RFID) chips embedded in drug packaging can leave a data "exhaust trail" that links back to information that health care providers would normally consider deeply confidential.[26]

Individuals may also offer up health data directly, further risking the generation of PII. For example, programs such as the "Blue Button" initiative allow patients to directly download their personal health records.[27] Once downloaded, many of these records lose the protections afforded to them by federal health privacy statutes such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009.[28]

---

[23] *See* Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331, 347–50 (2007) (discussing various means used to compile personal health information, including hackers, foreign data processes, public records, and consumer purchase information); *see also* Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism* 19–21 (Ind. Univ. Robert H. McKinney Sch. of Law Research Paper, Paper No. 2013-36, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2320088, *archived at* http://perma.cc/7LZT-CF2P (noting that this expansion in the health care industry has focused on making this information available to the individual, rather than to health care providers).

[24] *See Report Finds Correlation Between Health Data Breaches, Fraud Cases*, IHEALTHBEAT (Apr. 30, 2012), http://www.ihealthbeat.org/articles/2013/4/30/report-finds-correlation-between-health-data-breaches-fraud-cases.aspx, *archived at* http://perma.cc/S4AM-ZAKH.

[25] Terry, *supra* note 19, at 394.

[26] *Id.*

[27] *See Your Health Records: About Blue Button*, HEALTHIT.GOV http://www.healthit.gov/patients-families/blue-button/about-blue-button, *archived at* http://perma.cc/9S94-4C4P (last visited Dec. 2, 2013).

[28] *See* Terry, *supra* note 19, at 394. HIPAA enacted a regime of privacy and security regulations for the health care industry that were later augmented by the HITECH Act. *See* HIPAA, Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified as amended in scattered sections of 26, 29, and 42 U.S.C.); HITECH, Pub. L. No. 111-5, 123 Stat. 115, 226 (2009) (codified as amended in scattered sections of 42 U.S.C.); Thomas R. McLean & Alexander B. McLean, *Dependence on Cyberscribes—Issues in e-Security*, 8 J. BUS. & TECH. L. 59, 90–92 (2013); Jo-Ellyn Sakowitz Klein et al., *Health Sector Braces for Wide Impact of the New HITECH Omnibus Rule*, 25 INTELL. PROP. & TECH. L.J. 10, 10 (2013). Recently, an omnibus rule was enacted, overhauling this existing regime—and essentially synthesizing and strengthening these protections. Klein et al., *supra*, at 10; Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules, 78 Fed. Reg. 5566 (Jan. 25, 2013) (to be codified at 45 C.F.R. pts. 160, 164). This new rule sets out security standards by which health plans, health care clearinghouses, and health care providers

Other self-help health and fitness approaches, such as "Quantified Self" and "PatientsLikeMe," generate data sets that help identify or predict health attributes.[29] When these data sets are cross-referenced with traditional health information, as Big Data is designed to do, it is possible to generate a detailed picture about a person's health, including information the person may never have disclosed to a health care provider.[30] By combining the use of these data sets with predictive analytics, Big Data can dramatically increase the amount of related data that may be considered private.[31]

But these privacy problems go beyond just increasing the amount and scope of potentially private information. Based on existing publicly available information, Big Data's processes can generate a predictive model of what has a high probability of being PII, essentially *imagining* an individual's data. For example, in the *New York Times* article about Target predicting pregnancy, Target had never collected data showing that any particular female customer was pregnant—a fact that most people would almost assuredly consider to be very personal and intimate information.[32] Instead, Target predicted this information.[33] Furthermore, the prediction was just as personally sensitive as if it had been collected or shared inappropriately.[34] Target also used the predictive privacy information in a similarly personally sensitive manner by exploiting it for marketing purposes.[35] Nevertheless, because it did not collect the information from any first or third party, Target had no obligation under current privacy regimes to give notice to, or gather consent from its customers in the same way that direct collection protocols require.[36] In the context of health information, this is likely to lead to a

must comply. *Id.* at 5567. Such standards include administrative, physical, and technical safeguards to the electronic storage of protected health information. *Id.* at 5567–69.

[29] *See* Terry, *supra* note 23, at 19–21. "Quantified Self" is a program that allows patients to track their activity and other health inputs such as heart rate and oxygen levels to improve their lifestyles. *See Quantified Self: Self Knowledge Through Numbers*, QUANTIFIED SELF, http://www.quantifiedself.com, *archived at* http://perma.cc/VUZ2-XBJN (last visited Jan. 15, 2014). Similarly, "PatientsLikeMe" allows those living with health conditions such as cancer and diabetes to share information about their treatment and symptoms in order to aggregate information and provide suggestions for possible steps. *See About Us*, PATIENTSLIKEME, http://www.patientslikeme.com/about, *archived at* http://perma.cc/E35S-8AYZ (last visited Jan. 15, 2014).

[30] *Counting Every Moment*, ECONOMIST (May 3, 2012), http://www.economist.com/node/21548493, *archived at* http://perma.cc/YW74-Z44P.

[31] *See generally* Paul Schwartz & Dan Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814 (2001) (arguing for a more flexible approach to PII that tracks the "risk of identification" along a spectrum and attributing the shortcomings of the current PII approach to the increase of PII generated by Big Data).

[32] Duhigg, *supra* note 5, § 6 (Magazine), at 30.

[33] *Id.*

[34] *Id.*

[35] *Id.*

[36] *Id.*; *see infra* notes 75–78 and accompanying text (discussing current regulations of privacy information).

surge in the use of Big Data to create "surrogate" health data because of its less regulated nature.[37]

Moreover, the nature of Big Data's dynamic analytical tools is such that the privacy problems of predictive algorithms are often themselves unpredictable, and their effects may not even be fully understood by their programmers.[38] As computer scientists have shown, in many contexts, it is *impossible* to guarantee differential privacy when using a learning algorithm that draws data from a continuous distribution.[39] In other words, we cannot know in advance exactly when a learning algorithm will predict PII about an individual; therefore, we cannot predict where and when to assemble privacy protections around that data. When a pregnant teenager is shopping for vitamins, could she predict that any particular visit or purchase would trigger a retailer's algorithms to flag her as a pregnant customer? And at what point would it have been appropriate to give notice and request her consent?

So how exactly does one define this type of privacy problem? The following examples highlight the predictive privacy harms of Big Data.[40] Not only does Big Data's use have the potential to circumvent existing antidiscrimination regulations, but it may also lead to privacy breaches in health care and law enforcement.

1. Predictive Privacy and Discriminatory Practices

Predictive privacy harms can manifest as discriminatory practices that circumvent current regulations. For decades, there have been laws prohibiting various discriminatory practices. In the real estate industry, for instance, such legislation prohibits marketing that excludes renters and buyers who fall within racial, gender, or religious categories.[41] This legislation works

---

[37] Terry, *supra* note 19, at 391–92. Big Data's ability to synthesize previously available information from various datasets similarly threatens privacy by enabling the "re-identification" of personal information or identities that have been stripped away. *See generally* Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, 2008 IEEE SYMP. ON SEC. & PRIVACY 111 (demonstrating via an anonymous Netflix user database that users may be identified with as few as five personal attributes); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) (highlighting the ease of re-identifying anonymous datasets, discussing how this harms current privacy law, and suggesting solutions).

[38] Kamalika Chaudhuri & Daniel Hsu, *Sample Complexity Bounds for Differentially Private Learning*, 19 JMLR: WORKSHOP & CONF. PROC. 155, 155–56 (2011).

[39] *Id.*

[40] *See infra* notes 41–74 and accompanying text.

[41] *See* Fair Housing Act of 1968, 42 U.S.C. § 3604(c) (2006). The Fair Housing Act of 1968 prohibits the making, printing, or publication of any "notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin." *Id.* Further-

because there are certain moments in the life cycle of housing advertisements that can act as flashpoints for enforcement.[42] Enforcers can monitor published housing advertisements by particular landlords in specific media and check for use of explicit language indicative of discriminatory intent, such as "female renters preferred."[43] Enforcement cases involve presenting the text of the advertisement to the adjudicator along with other evidence of intent.[44]

The use of Big Data may allow landlords and real estate companies to shift away from general advertising in media outlets and circumvent anti-discrimination enforcement mechanisms by isolating correlative attributes that they can use as a proxy for traits such as race or gender. Such predictive practices already occur in other industries with the power of Big Data.[45] This can be partly attributed to Big Data's ability to generate a detailed picture of individuals with even discrete online activity, such as "liking" things on Facebook.[46] For example, in the credit loan industry, federal regulations prohibit discrimination in access to credit.[47] Despite these regulations, com-

---

more, § 3604(a) prohibits "otherwise mak[ing] unavailable or deny[ing], a dwelling to any person because of race, color, religion, sex, familial status, or national origin." *Id.* § 3604(a). Whether using Big Data to generate profiles that de facto discriminate under § 3604(a) would violate § 3604(c) is unknown and presumably unlikely, given the intent requirement. *See id.* § 3604(c).

[42] *See* Stephen Collins, Comment, *Saving Fair Housing on the Internet: The Case for Amending the Communications Decency Act*, 102 NW. U. L. REV. 1471, 1490–93 (highlighting the decline in § 3604(c)'s effectiveness in preventing housing discrimination and attributing this trend to the rise of new technologies for advertising).

[43] *See generally* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008) (noting that a roommate search service's requirement for users to disclose sex, sexual orientation, and family status may be discriminatory and violative of the Fair Housing Act); Chi. Lawyers' Comm. for Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666 (7th Cir. 2008) (noting that the publication of housing advertisements prohibiting minorities and children may be discriminatory and violative of the Fair Housing Act).

[44] *See generally, e.g.*, Hous. Opportunities Made Equal, Inc. v. Cincinnati Enquirer, Inc., 943 F.2d 644 (6th Cir. 1991) (considering evidence of advertisements that featured predominantly white models as discriminatory and liable under § 3604(c)); Ragin v. N.Y. Times Co., 923 F.2d 995 (2d Cir. 1991) (same); United States v. Hunter, 459 F.2d 205 (4th Cir. 1972) (considering evidence of newspaper advertisements that described rental as a "white home" as discriminatory and liable under § 3604(c)).

[45] *See generally* JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH (2012) (discussing how media, including advertisements and entertainment, uses Big Data to acquire and create individual profiles for consumers).

[46] *See generally* Woodrow Hartzog & Evan Selinger, *Big Data in Small Hands*, 66 STAN. L. REV. ONLINE 81 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/big-data-small-hands, *archived at* http://perma.cc/6B9P-VGTL (highlighting Big Data's power to generate highly detailed individual profiles with little social media information); Michael Kosinski, et al., *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802 (2013) (finding that "highly sensitive personal attributes" could be predicted with high degrees of success from "Facebook Likes").

[47] *See* Equal Credit Opportunity Act (ECOA), 15 U.S.C. §§ 1691–1691f (2012).

panies have utilized Big Data models to identify and avoid Internet users with low credit scores when posting advertisements for loans.[48]

Accordingly, Big Data may eliminate housing suppliers' need to disclose their potentially discriminatory preferences in their advertisements. Instead, the housing providers could design an algorithm to predict the relevant PII of potential buyers or renters and advertise the properties only to those who fit these profiles. In the housing market, providers may adopt these practices and increasingly rely on publicly available personal information to generate these profiles. Just as Big Data may be used to prevent candidates from seeing loans that might be advantageous to them, housing suppliers could potentially use Big Data to discriminate, all while circumventing the fair housing laws.[49]

Big Data's ability to discriminate while maneuvering around privacy regulations comes from its methodology. Not only can massive amounts of online behavior be collected and assessed to compute the probabilities of an individual's particular demographic characteristic, but that predictive analysis can also become a form of PII itself.[50] Moreover, this process can predict highly intimate information, even if none of the individual pieces of data could be defined as PII.[51] Although these predictive processes may generate an inaccurate characterization, such processes nevertheless create a model of possible personal information and associate it with an individual.[52] Accordingly, harms can result regardless of the model's accuracy.

---

[48] Michael Fertik, *The Rich See a Different Internet Than the Poor*, SCI. AM., (Feb. 18, 2013), http://www.scientificamerican.com/article.cfm?id=rich-see-different-internet-than-the-poor, *archived at* http://perma.cc/3PCW-GBDB (stating that if Big Data analysis indicates a poor credit record for the user, "you won't even see a credit offer from leading lending institutions, and you won't realize that loans are available to help you with your current personal or professional priorities").

[49] *Cf.* Alistair Croll, *Big Data Is Our Generation's Civil Rights Issue, and We Don't Know It*, SOLVE FOR INTERESTING (July 31, 2012, 12:40 PM), http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/, *archived at* http://perma.cc/D3T-QCW9 (illustrating how Big Data can be used for discriminatory purposes); Alistair Croll, *Followup on Big Data and Civil Rights*, SOLVE FOR INTERESTING (Aug. 28, 2012, 11:29 AM), http://solveforinteresting.com/followup-on-big-data-and-civil-rights/, *archived at* http://perma.cc/ZP9K-NGFC (same).

[50] *See* Tene & Polonetsky, *supra* note 14, at 256 (suggesting that Target had generated sensitive personal information by using Big Data's predictive analytics when it determined that a teen was pregnant).

[51] *See* Schwartz & Solove, *supra* note 31, at 1841–45 (explaining how Big Data can transfer previously anonymous data into PII through re-identification).

[52] *See* Cynthia Dwork & Deirdre Mulligan, *It's Not Privacy, and It's Not Fair*, 66 STAN. L. REV. ONLINE 35, 36–38 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/its-not-privacy-and-its-not-fair, *archived at* http://perma.cc/35X-E9XS; Ian Kerr & Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, 66 STAN. L. REV. ONLINE 65, 69 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/prediction-preemption-presumption, *archived at* http://perma.cc/CXU5-54V2.

2. Health Analytics and Personalized Medicine

As one noted health law scholar has written, many consider technology pivotal to addressing issues in health care because technology's rapid progress has facilitated affordable and easy access to research advancements and patient records.[53] Given these benefits, the use of Big Data in health care seems particularly attractive.[54] Some even predict that it will bring forth "personalized medicine" —analyzing large data sets of patient information to make diagnostic predictions and treatment suggestions.[55] Nonetheless, similar issues to those previously discussed arise when Big Data is used to address health problems.

For example, Big Data's methodology may harm the privacy of health information. Such personalized models would require Big Data practices to access very detailed personal health information about an individual as well as thousands of others' profiles for comparison.[56] The generated predictions and treatment suggestions would be associated with the individual in the same way

---

[53] Terry, *supra* note 19, at 385. Terry notes:

> Technology, not surprisingly, is viewed as holding the solution [to rising health care costs] because "[a]dvances have made vast computational power affordable and widely available, while improvements in connectivity have allowed information to be accessible in real time virtually anywhere" affording "the potential to improve health care by increasing the reach of research knowledge, providing access to clinical records when and where needed, and assisting patients and providers in managing chronic diseases."

Id. (alteration in original) (quoting INST. OF MED., BEST CARE AT LOWER COST: THE PATH TO CONTINUOUSLY LEARNING HEALTH CARE IN AMERICA 112 (Mark Smith et al. eds., 2013)).

[54] *Id.*

[55] Narges Bani Asadi, *The Personalized Medicine Revolution Is Almost Here*, VENTUREBEAT (Jan. 27, 2013, 12:23 PM), http://venturebeat.com/2013/01/27/the-personalized-medicine-revolution-is-almost-here/, *archived at* http://perma.cc/LJP4-ZL75; Press Release, Dep't for Bus., Innovation, & Skills and Prime Minister's Office, *£30 Million Investment in Health Research Centre to Tackle Major Diseases* (May 3, 2013), https://www.gov.uk/government/news/30-million-investment-in-health-research-centre-to-tackle-major-diseases (expressing the Prime Minister's belief in Big Data's promise for health care); *see also* Terry, *supra* note 19, at 394 ("It will not be long until patient level information is combined with large existing data sets [that] will generate far more accurate predictive modeling, personalization of care, assessment of quality and value for many more conditions, and help providers better manage population health and risk-based reimbursement approaches." (quoting Robert Kocher & Bryan Roberts, *Meaningful Use of Health IT Stage 2: The Broader Meaning*, HEALTH AFF. BLOG (Mar. 15, 2012, 1:24 PM), http://healthaffairs.org/blog/2012/03/15/meaningful-use-of-health-it-stage-2-the-broader-meaning/, *archived at* http://perma.cc/8JP7-K7U8).

[56] Sarah A. Downey, *How to Use 23andMe Without Giving Up Your Genetic Privacy*, VENTUREBEAT (Sept. 20, 2013, 10:19 AM), http://venturebeat.com/2013/09/20/how-to-use-23andme-without-giving-up-your-genetic-privacy/, *archived at* http://perma.cc/367S-ZZHB. *See generally* 23ANDME, http://www.23andme.com, *archived at* http://perma.cc/MF5Z-4B2H (last visited Dec. 2, 2013) (providing an example of a genetic testing company that relies on Big Data methodology); PERSONAL GENOME PROJECT, http://www.personalgenomes.org, *archived at* http://perma.cc/C8SF-4AFW (last visited Dec. 2, 2013) (providing an example of a large-scale personal genomics study that relies on Big Data methodology).

as PII. As one noted health scholar emphasizes, HIPAA/HITECH's security and privacy standards for electronic health records apply to entities of health plans, health care clearinghouses, and health care providers; in contrast, it is unclear whether these regulations will apply to organizations that are not so characterized, but who still receive personal health information from individuals or by generating it through Big Data.[57] Thus, even health information—one of the most highly protected types of personal information—will be increasingly vulnerable in the context of Big Data and predictive analytics.

3. Predictive Policing

Law enforcement agencies throughout the United States are turning to predictive policing models of Big Data in the hopes that they will shine investigative light on unsolved cases or help prevent future crimes.[58] This model uses the date, time, type, and location of recent crimes and combines that data with historical crime data to identify "hot spots" that become the focus of officer patrols.[59]

Big Data's ability to analyze large amounts of data may lead to predictive privacy harms for individuals targeted by law enforcement. With Big Data, it takes very little to connect time, place, and location with individuals, especially when combined with other data sets.[60] Moreover, the predictions that these policing algorithms make—that particular geographic areas are more likely to have crime—will surely produce more arrests in those areas by directing police to patrol them.[61] This, in turn, will generate more

---

[57] Terry, *supra* note 19, at 386 (questioning the value of HIPAA/HITECH protections, which are "designed to keep unauthorized data aggregators out of our medical records," when Big Data "allows the creation of surrogate profiles of our medical selves"); *see* Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules, 78 Fed. Reg. 5566 (Jan. 25, 2013) (to be codified at 45 C.F.R. pts. 160, 164). In addition to circumventing protected domains entirely, Big Data may also benefit from one of a number of carve-outs to traditionally protected HIPAA/HITECH domains. Terry, *supra* note 19, at 408. For example, the Big Data task of "running data analytics against a hospital's [Electronic Medical Records] data" in order to "look[] for disease predictors" may be categorized as a "quality improvement under 'health care operations,'" and therefore be exempt from regulation. *Id.*

[58] *See* Zach Friend, *Predictive Policing: Using Technology to Reduce Crime*, FBI L. ENFORCEMENT BULL. (Apr. 9, 2013), http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/2013/April/predictive-policing-using-technology-to-reduce-crime, *archived at* http://perma.cc/D8GC-2EDC.

[59] *Id.*

[60] *Cf.* Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, SCI. REP., Mar. 25, 2013 at 1, 1, http://www.nature.com/srep/2013/130325/srep01376/pdf/srep01376.pdf, *archived at* http://perma.cc/BUS5-26VS (explaining how Big Data can re-identify mobile phone data to track individuals). Existing databases such as historical crime data enhance Big Data's ability to connect individuals with the available information. Andrew G. Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259, 317 (2012).

[61] Ferguson, *supra* note 60, at 265–85 (explaining predictive policing models).

"historical crime data" for those areas and increase the likelihood of patrols.[62] For those who live there, these "hot spots" may well become as much PII as other demographic information.[63]

Law enforcement's use of Big Data in other ways may similarly lead to abuse. Similar analytics are used in "fusion centers," —information hubs created by the U.S. Department of Homeland Security and the U.S. Department of Justice to share personal data held by such agencies as the Central Intelligence Agency, Federal Bureau of Investigation, and the military.[64] This aggregation of various agencies' data allows law enforcement to predict or flag individuals as suspicious or worthy of investigation, search, or detention based on the agency's outlined criteria.[65] As two scholars note, this method may sometimes lead to erroneous results:

> In one case, Maryland state police exploited their access to fusion centers to conduct surveillance of human rights groups, peace activists, and death penalty opponents over a nineteen-month period. Fifty-three political activists eventually were classified as "terrorists," including two Catholic nuns and a Democratic candidate for local office. The fusion center shared these erroneous terrorist classifications with federal drug enforcement, law enforcement databases, and the National Security Administration, all without affording the innocent targets any opportunity to know, much less correct, the record.[66]

When combined and constructed into a composite prediction of a person, such analytics have very serious consequences for personal privacy. For example, in 2012, in the U.S. Supreme Court decision in *United States v. Jones*, Justice Sonia Sotomayor's concurring opinion expressed serious concerns about invasions of privacy that could result from direct collection of massive amounts of personal data—specifically, the government's ability "to assemble data that reveal private aspects of identity," through means such as Global Position System (GPS) monitoring.[67] Such unrestrained governmental power may be susceptible to abuse, endanger individuals' right to privacy, and weaken individuals' trust in government.

With predictive policing, the government is not only *watching and collecting* massive amounts of information about individuals, but it is also using predictive analytics to *generate* "data that reveal private aspects of iden-

---

[62] *See id.*

[63] *See id*.

[64] David Gray & Danielle Citron, *The Right to Quantitative Privacy*, 98 MINN. L. REV. 62, 67 (2013).

[65] *Id*.

[66] *Id.* at 81 (footnotes omitted).

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

tity," which is subject to abuse in similar ways.[68] Even the companies designing software for law enforcement reveal that their systems do not necessarily conform to practical or ethical standards.[69] As one of these companies' former privacy officers has stated, "[G]eeks like me can do stuff like this, we can make stuff work—it's not our job to figure out if it's right or not. We often don't know."[70] This software can have particularly harmful impacts on racial profiling and other avenues of discrimination. Furthermore, it may circumvent the goal of programs designed to promote rehabilitation and reincorporation through "Clean Slate" laws.[71] Whereas these laws allow certain nonviolent offenders to expunge their criminal records to gain better education, employment, and housing opportunities, law enforcement's use of data analytics may resurface these prior convictions.[72]

Furthermore, Big Data will likely exacerbate current concerns of protecting Fourth Amendment rights in light of new technologies. Recently, questions about the constitutional limits on public data surveillance, such as the GPS tracking at issue in *Jones*, continue to test the courts' interpretation of Fourth Amendment precedents.[73] The generative data-making practices of Big Data will only place further strain on these issues because its approaches to policing and intelligence may be both qualitatively *and* quantitatively different from the surveillance approaches addressed in *Jones*.[74]

---

[68] *Cf. id.* (suggesting that government collection of massive amounts of personal data is subject to abuse); Kerr & Earle, *supra* note 52, at 69 (highlighting organizations' and the government's use of existing personal data and predictive algorithms to extrapolate conclusions unrelated to that data). *See generally* Patricia L. Bellia, *The Memory Gap in Surveillance Law*, 75 U. CHI. L. REV. 137 (2008) (describing the various methods through which the government may collect data about individuals).

[69] Jordan Robertson, *How Big Data Could Help Identify the Next Felon—or Blame the Wrong Guy*, BLOOMBERG (Aug. 15, 2013, 12:01 AM), http://www.bloomberg.com/news/2013-08-14/how-big-data-could-help-identify-the-next-felon-or-blame-the-wrong-guy.html, *archived at* http://perma.cc/HEQ8-5PNT.

[70] *Id*. This privacy officer had worked at Intelius, a company that designs software to predict whether an individual will be a felon with minimal data and considerable predictive guesswork. *Id.* Although the software can make accurate predictions, it has also generated some false positives. *Id.*

[71] *See* Second Chance Act of 2007, 42 U.S.C. § 17501 (Supp. V 2011) (seeking to reduce recidivism rates by providing employment, housing, and other assistance to non-violent criminal offenders).

[72] *See* Lahny R. Silva, *Clean Slate: Expanding Expungements and Pardons for Non-Violent Federal Offenders*, 79 U. CIN. L. REV. 155, 164–74, 198–99 (2011) (stating that criminal convictions bars individuals from various opportunities and that legislation designed to expunge their records may decrease recidivism).

[73] *See, e.g.*, *Jones*, 132 S. Ct. at 956; United States v. Katzin, 732 F.3d 187, 193–94 (3d Cir. 2013) (considering whether GPS data may be included as evidence if authorities obtained the GPS without a warrant).

[74] Gray & Citron, *supra* note 64, at 112–24.

C. *Predictive Privacy Harms Threaten to Marginalize Traditional Privacy Protections*

In light of these predictive privacy harms, it is worth considering what an appropriate set of privacy protections might be to address them. Traditionally, American civil privacy protections have focused on regulating three main activities: information collection, processing, and disclosure.[75] For example, the Electronic Communications Privacy Act of 1986 (ECPA) prohibits the unauthorized collection of communications content;[76] the Fair Credit Reporting Act prohibits the use of financial records for certain purposes;[77] and the Video Privacy Protection Act of 1988 prohibits the disclosure of video rental records.[78]

Big Data has the potential to elude all three of these approaches primarily because of the unpredictable nature of its predictive privacy harms. From the perspective of data collection regulations, one cannot assess the predictive privacy risks from the collection of a single data point such as a single tweet, a single "like" on Facebook, a single web search, or a single afternoon drive recorded on a GPS.[79] Regulating collection is hard, and regulating predictive analytics is even more difficult. Nor can one necessarily predict when a certain form of information processing will produce predictive privacy harms. Even disclosure regulations become complicated because Big Data systems do not create PII at the point of collection. Oftentimes, the data that ends up being personally identifying may not yet exist during the most significant data transfers. This is seen in predictive policing systems, where numerous data collections and transfers can occur before any predictive private harm comes into existence. In fact, it may only be after all transfers are complete that the predictions occur. For example, if the FBI collected crime data from numerous local law enforcement databases to predict areas likely to house sex offenders, it would be impossible to identify the resulting harm when the data transfers occurred. After all, Big Data promises to accomplish such previously unattainable tasks. Thus, unless one decides that privacy regulations must govern all data ever collected, processed, or disclosed, deciding where and when to draw lines around these activities becomes extremely difficult with respect to Big Data information practices.

---

[75] Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 484–552 (2006) (detailing these three categories).

[76] *See* 18 U.S.C. §§ 2510–2522, 2701–2711, 3121–3127 (2012).

[77] 15 U.S.C. §§ 1681–1681t (2012).

[78] 18 U.S.C. § 2710 (2012), *amended by* Video Privacy Protection Act Amendments Act of 2012, Pub. L. No. 112-258, 126 Stat. 2414 (2013).

[79] *See* Terry, *supra* note 14, at 257–63 (describing the difficulties in applying privacy law to Big Data); Schwartz & Solove, *supra* note 31, at 1845–47 (describing the difficulties of characterizing Big Data results as PII).

Even the anchoring concept of most modern privacy regulations—PII—may fail to provide a sufficient nexus for future Big Data privacy regulation. Big Data's analytics are simply too dynamic and unpredictable to determine if and when particular information or analyses will become or generate PII.[80] Instead, one may only observe the problem in hindsight, such as with predictive policing or in the Target pregnancy case.[81]

Moreover, predictive privacy harms may marginalize the broader frameworks for privacy. For decades, privacy policymakers have relied on a set of Fair Information Practice Principles ("FIPPs") as guidelines for adapting existing privacy laws and developing new ones, especially in light of new technological developments or information practices.[82] Various versions of the FIPPs exist, but in general, the Federal Trade Commission (FTC) suggests these core principles: (1) notice/awareness; (2) choice/consent; (3) access/participation; (4) integrity/security; and (5) enforcement/redress.[83] In February 2012, the White House also outlined and released its own FIPPs, which included the principles of individual control, transparency, respect for context, security, access and accuracy, focused collection, and accountability.[84]

---

[80] Schwartz & Solove, *supra* note 31, at 1845–47.

[81] *See* Duhigg, *supra* note 5, § 6 (Magazine), at 30 (discussing Target's use of Big Data); *supra* notes 58–74 and accompanying text (describing predictive policing).

[82] *See* ROBERT GELLMAN, FAIR INFORMATION PRACTICES: A BASIC HISTORY 1 (2013), http://bobgellman.com/rg-docs/rg-FIPShistory.pdf, *archived at* http://perma.cc/ZBY-5LDP.

[83] FED. TRADE COMM'N, PRIVACY ONLINE: A REPORT TO CONGRESS 7–11 (1998), *available at* http://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/priv-23a.pdf, *archived at* http://perma.cc/PC3T-XMQQ; *see* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE 11 (2012), *available at* http://www.ftc.gov/os/2012/03/120326privacyreport.pdf, *archived at* http://perma.cc/DE92-6R73.

[84] THE WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 1 (2012), *available at* http://www.whitehouse.gov/sites/default/files/privacy-final.pdf, *archived at* http://perma.cc/6ZCR-742V. The White House explains its FIPPs in its "Consumer Privacy Bill of Rights":

> — Individual Control: Consumers have a right to exercise control over what personal data companies collect from them and how they use it.
> — Transparency: Consumers have a right to easily understandable and accessible information about privacy and security practices.
> — Respect for Context: Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.
> — Security: Consumers have a right to secure and responsible handling of personal data.
> — Access and Accuracy: Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate.
> — Focused Collection: Consumers have a right to reasonable limits on the personal data that companies collect and retain.

Although the FTC and the White House share some FIPPs, the White House's version includes additional approaches to information regulation that attempt to expand the scope of FIPPs, especially in terms of control over the information at issue and with more focus on the data's user and categorization. Yet even these broadened principles depend on not only knowing which information is considered personal data but also providing notice, choice, and control to users ex ante any privacy harm. Privacy law is primarily concerned with causality, whereas Big Data is generally a tool of correlation.[85] This contrast makes FIPPs-style approaches to privacy protection particularly difficult with respect to Big Data.

But how does one give notice and get consent for innumerable and perhaps even yet-to-be-determined queries that one might run that create "personal data"? How does one provide consumers with individual control, context, and accountability over such processes? Such difficulties suggest that frameworks like FIPPs may fail to regulate predictive privacy harms because they focus on data collection and retention while using notice-and-consent models.[86]

As Big Data is versatile, dynamic and unpredictable, traditional notions of privacy that isolate certain categories of information—such as PII—to regulate collection, utilization, or disclosure are ill-suited to address these emerging risks.[87] Even omnibus privacy approaches, such as the European Union's "right to be forgotten," will likely struggle with Big Data's ability to recall or even reconstruct an individual's personal information based on past or present data.[88] These types of privacy problems demand a

---

— Accountability: Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.

Id.

[85] *See, e.g.*, Rubinstein, *supra* note 3, at 76 (stating that Big Data relies on correlation instead of causation); Gil Press, *Big Data News Roundup: Correlation vs. Causation*, FORBES (Apr. 19, 2013, 10:23 AM), http://www.forbes.com/sites/gilpress/2013/04/19/big-data-news-roundup-correlation-vs-causation/, *archived at* http://perma.cc/5QJQ-LDAC (summarizing various media reports that highlight Big Data's use of correlation).

[86] *See* Dwork & Mulligan, *supra* note 52, at 36–38 (explaining why current privacy frameworks do not adequately address Big Data); Hartzog & Selinger, *supra* note 46, at 82–83 (stating that FIPPs do not adequately address Big Data).

[87] *See* Dwork & Mulligan, *supra* note 52, at 36–38; Natasha Singer, *Acxiom Lets Consumers See Data It Collects*, N.Y. TIMES, Sept. 5, 2013, at B6. For example, one notorious data broker, Acxiom, now lets customers see and change the data it collects about them individually. Singer, *supra*. Acxiom, however, does not allow them to change the analytics it uses to assess the data for sale to marketers. *Id.* This is a sign that transparency and regulation of individual data collection is not likely to serve as an effective gatekeeping function for controlling privacy harms. *See id.*

[88] *See, e.g.*, Meg L. Ambrose, *It's About Time: Privacy, Information Life Cycles, and the Right to Be Forgotten*, 16 STAN. TECH. L. REV. 369, 385–87 (2013) (highlighting concerns about the right to be forgotten); Jeffrey Rosen, *The Right to Be Forgotten*, 64 STAN. L. REV. ONLINE 88, 90–92,

shift in thinking that can approach the problem with the same dynamic and flexible capacities that Big Data itself provides.

II. WHY BIG DATA NEEDS PROCEDURAL DUE PROCESS

> *The heart of the matter is that democracy implies respect for the elementary rights of men, however suspect or unworthy; a democratic government must therefore practice fairness; and fairness can rarely be obtained by secret, one-sided determination of facts decisive of rights.*
>
> —J. Frankfurter[89]

As noted in Part I, the power of Big Data analyses to evade or marginalize traditional privacy protections and frameworks, its drive to bring visibility to the invisible, and its dynamic and unpredictable nature all present challenges to thinking about how privacy and Big Data can coexist. In response, this Article proposes an alternative approach—procedural data due process. Rather than attempt regulation of personal data collection, use, or disclosure ex ante, procedural data due process would regulate the fairness of Big Data's analytical processes with regard to how they use personal data (or metadata derived from or associated with personal data) in any adjudicative process, including processes whereby Big Data is being used to determine attributes or categories for an individual. For example, if a health insurance provider used Big Data to determine the likelihood that a customer has a certain disease and thus denied coverage on that basis, the customer would have a data due process right with regard to that determination. Similarly, if a potential employer used Big Data to predict how honest certain job applicants might be, these applicants could then exercise their data due process rights.[90]

What would such a regulatory process entail? This Part describes some of the history of due process in Anglo-American law. This Part then demonstrates why due process's embedded values and its traditional procedures in

---

http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-88.pdf, *archived at* http://perma.cc/X8HL-EP69 (discussing shortcomings of the European Union's right to be forgotten). As described by the European Commission in its draft General Data Protection Regulation, the "right to be forgotten" mandates that, upon the individual's request, entities that collect or store data must delete data pertaining to that individual. *Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*, at 51–53, COM (2012) 11 final (Jan. 25, 2012), *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:EN:PDF, *archived at* http://perma.cc/CA2F-L253.

[89] Joint Anti-Fascist Refugee Comm. v. McGrath, 341 U.S. 123, 170 (1951) (Frankfurter, J., concurring).

[90] *See* Steve Lohr, *Big Data, Trying to Build Better Workers*, N.Y. TIMES, Apr. 21, 2013, at BU4 (illustrating how Big Data is used by employers to identify ideal traits in job applicants).

courts of law and administrative proceedings may be well-suited for an analogous system regulating private use of Big Data to mitigate predictive privacy harms. Finally, this Part discusses what procedural data due process would involve and some possible implementations.

A. *The Historical Role of Due Process*

Throughout the history of Anglo-American law, procedural due process has served as a set of constraints on adjudication—the process of deciding.[91] Adjudications are arguably similar to the type of models and determinations that predictive algorithms create based on massive data sets. Just as information drives Big Data determinations, so does it drive litigation, legal strategies, and legal outcomes.[92] Law—much like computer code and data—has its own information rules that are governed by various frameworks, from formal rules like the Federal Rules of Civil Procedure to common law and constitutional doctrines such as due process.[93]

Our modern conception of due process is derived from two foundational sources. First is the Magna Carta, which understood due process to mean that "[n]o freemen shall be taken or {and} imprisoned or disseised or exiled or in any way destroyed, nor will we go upon him nor send upon him, except by the lawful judgment of his peers or {and} by the law of the land."[94] Due process then made its way into the U.S. Constitution as part of the Fifth Amendment, which states: "No person shall . . . be deprived of life, liberty, or property, without due process of law."[95]

There are two important components to note here. The first is the prohibition on deprivation. The subjects—life, liberty, and property—are obviously each broad categories that have, at times, defined the core components of citizenship.[96] They represent qualitatively the level of seriousness that the deprivation must constitute in order to invoke due process protec-

---

[91] Ryan C. Williams, *The One and Only Substantive Due Process Clause*, 120 YALE L.J. 408, 419–21 (2010).

[92] *See* Fredric M. Bloom, *Information Lost and Found*, 100 CALIF. L. REV. 636, 636 (2012).

[93] *Id.* at 640; *see also* LAWRENCE LESSIG, CODE: VERSION 2.0, at 4–8 (2006) (discussing code's role in cyberspace as law and exploring interpretations for lawyers and citizens).

[94] MAGNA CARTA c. 39 (1215), *reprinted in* THE GREAT CHARTER: FOUR ESSAYS ON MAGNA CARTA AND THE HISTORY OF OUR LIBERTY 132 (1965); *see also* Hurtado v. California, 110 U.S. 516, 521–25 (1884) (explaining the Magna Carta's interpretation of due process and differentiating it from the Fifth Amendment); Murray's Lessee v. Hoboken Land & Improvement Co., 59 U.S. (18 How.) 272, 276–78 (1856) (analogizing "due process of law" to the Magna Carta's reference to "the law of the land").

[95] U.S. CONST. amend. V. This provision focused on federal state actions. *Id.* The Fourteenth Amendment also contains a similar clause that extends due process protections to individual state actions. *Id*. amend. XIV.

[96] *See* Williams, *supra* note 91, at 420–22 (noting that each interpretation of due process includes a focus on life, liberty, and property).

tion and reflect the type of harm that we wish to prevent.[97] The category of liberty is especially important to consider in the context of privacy and predictive privacy harms. Both John Locke and William Blackstone described liberty as an individual's unabridged natural right to follow his own will.[98] If one considers privacy to be "the right to be let alone" and to have some freedom of self-determination and autonomy, then it fits well within the liberty category.[99] Property and other interests are also implicated, especially as Big Data analytics are integrated into decisions concerning housing opportunities, employment, and credit provisioning.[100] Thus, predictive privacy harms seem well-suited for due process protection in terms of the type of subject matter covered.

The second component— "without due process of law" —is a means to enforce the probation: a process. But what constitutes this process? What are the underlying values that drive it? How would they fare as a means of regulating Big Data?

B. *Procedural Due Process in the Courts*

Today, procedural due process generally describes the constitutional requirement that any government deprivation of a liberty or property right must be preceded—at a minimum—by notice and the opportunity for a hearing on the matter before an impartial adjudicator.[101]

Historically, this conception of due process comes mainly from two landmark U.S. Supreme Court cases: *Mathews v. Eldridge*[102] and *Goldberg*

---

[97] *Id.*

[98] *See* 1 WILLIAM BLACKSTONE, COMMENTARIES *91 (stating that the right to personal liberty is "strictly natural" and "cannot ever be abridged . . . without the explicit permission of the laws"); JOHN LOCKE, TWO TREATISES OF GOVERNMENT 284 (Peter Laslett ed., Cambridge Univ. Press 1988) (1690) (describing personal liberty as "a liberty to follow [one's] own Will in all things" and "to be under no other restraint but the law of nature").

[99] *See* Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 387 (2008) (arguing that privacy "safeguards the integrity of our intellectual activities by shielding them from the unwanted gaze or interference of others"); Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARVARD L. REV. 193, 193 (1890) (defining privacy as "the right to be let alone").

[100] *See* Gray & Citron, *supra* note 64, at 100 (suggesting there is a right to information privacy based on substantive due process because continuous data collection harms individuals' self development and threatens their fundamental rights). These scholars also note that "the insidious, far-reaching and indiscriminate nature of electronic surveillance—and, most important, its capacity to choke off free human discourse that is the hallmark of an open society—makes it almost, although not quite, as destructive of liberty as 'the kicked-in-door.'" *Id.* (quoting Anthony Amsterdam, *Perspectives on the Fourth Amendment*, 58 MINN. L. REV. 349, 388 (1974)).

[101] *See* Hamdi v. Rumsfeld, 542 U.S. 507, 533 (2004); Cleveland Bd. of Educ. v. Loudermill, 470 U.S. 532, 542 (1985) ("An essential principle of due process is that a deprivation of life, liberty, or property 'be preceded by notice and opportunity for hearing appropriate to the nature of the case.'" (quoting Mullane v. Cent. Hanover Bank & Trust Co., 339 U.S. 306, 313 (1950))).

[102] 424 U.S. 319, 323–26, 333–35 (1976).

v. Kelly.[103] In 1970, in *Goldberg*, the Court held that procedural due process required an evidentiary hearing before the government could deprive a person of welfare benefits.[104] There, the New York City Department of Social Services allowed city caseworkers to terminate payments to welfare recipients whom they deemed ineligible. After the City terminated the recipients' welfare payments, the recipients could request a post-termination hearing challenging the decision.[105] The Court nonetheless concluded that this procedure inadequately protected the welfare recipients' procedural due process rights under the Fourteenth Amendment.[106] Writing for the majority, Justice William J. Brennan, Jr. explained the necessity for greater protection: "For qualified recipients, welfare provides the means to obtain essential food, clothing, housing, and medical care."[107] The Court therefore characterized the revocation of a welfare benefit as a governmentally sanctioned "grievous loss," which required the government to afford the individual certain procedural protections of due process before this loss.[108] The state need not resort to a complete judicial or quasi-judicial trial in order to satisfy due process.[109] Instead, the state must—at minimum—provide the potentially aggrieved party with the opportunity to be heard at a meaningful time and in a meaningful manner; adequate notice; the opportunity to present witnesses; and the ability to present arguments and evidence.[110] Thus, *Goldberg* set a fairly high procedural bar for any action that could deprive an individual of a property or liberty interest.[111]

In 1976, in *Mathews*, the Court retreated somewhat from this position when it addressed the Social Security Act's policy for the termination of disability benefits.[112] In many ways, this case was similar to *Goldberg*: in both cases, the state had deprived an individual of some government benefit

---

[103] 397 U.S. 254, 267–68 (1970).
[104] *Id.*
[105] *Id*. at 257–60.
[106] *Id.* at 263–64.
[107] *Id.* at 264.
[108] *Id*. at 263 (quoting *McGrath*, 341 U.S. at 168).
[109] *Id.* at 266.
[110] *Id*. at 267–68.
[111] *See id.*
[112] 424 U.S. at 323–26, 333–35 (differentiating its due process requirements in *Goldberg* from other decisions based on *Goldberg*'s facts). According to the Social Security Act, disabled workers bore the burden to prove their entitlement to benefits by showing that they were unable to perform their previous work—or any other gainful employment—because of a disability. *Id.* at 323–26. Local state agencies would review the evidence provided and make continuing determinations as to the worker's eligibility for aid. *Id.* If the agency felt that an aid recipient no longer qualified for disability relief, it would inform the recipient and the Social Security Administration (SSA) and provide both with a summary of the relevant evidence. *Id.* The SSA would then make a final determination; if the SSA terminated disability benefits, the recipient had the opportunity for a thorough review hearing. *Id.*

without the opportunity for a pre-termination hearing.[113] But the Court in *Mathews* concluded that the termination of disability payments did not require the same pre-termination hearing as the termination of welfare payments: "The private interest that will be adversely affected by an erroneous termination of benefits is likely to be less in the case of a disabled worker than in the case of a welfare recipient."[114] The Court relied upon its perceived difference in the financial burden of termination between disability aid recipients and welfare recipients.[115] Furthermore, the countervailing state interest in fiscal prudence and efficiency outweighed the potential for erroneous and harmful deprivation, thereby making additional procedural protections constitutionally unnecessary.[116]

To soften the supposedly stringent requirements of *Goldberg*, the Court in *Mathews* established a test for determining what courts must consider when judging the constitutionality of the deprivation.[117] This test consisted of balancing three factors:

> First, the private interest that will be affected by the official action; second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and finally, the Government's interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.[118]

Although *Mathews* shows that the level of due process required differs according to the gravity of the deprivation and the magnitude of the countervailing state interest, most cases over time have established four distinct procedural requirements that apply when the state deprives an individual of a state interest. Those requirements include (1) participatory procedures (i.e., the affected party is present); (2) a neutral arbiter; (3) prior process (i.e., the hearing precedes the adverse action); and (4) continuity (i.e., the hearing rights attach at all stages).[119] As this Article later discusses, these elements may be useful to address Big Data's privacy harms.

---

[113] *Compare id.* (allowing the disability aid recipient to appeal the SSA's termination of aid only after the SSA had made a final determination), *with Goldberg*, 397 U.S. at 257–60 (allowing the welfare recipient to appeal SSA's termination of payments only after the agency had made a final determination).

[114] 424 U.S. at 334–35.

[115] *Id.* at 342 (noting that despite the significant hardship that termination may impose on a disability recipient, the hardship is more significant for welfare recipients).

[116] *See id.* at 339–43.

[117] *Id.* at 335.

[118] *Id.*

[119] *See, e.g.*, *Hamdi*, 542 U.S. at 533; *Loudermill*, 470 U.S. at 542.

1. Eleven Elements of a Due Process "Hearing"

In addition to Supreme Court precedent, another valuable source for identifying elements of procedural due process is the seminal 1971 article, *Some Kind of Hearing*, by Judge Henry Friendly.[120] Similar to the balancing test in *Mathews*, Judge Friendly emphasizes that there is no specific checklist of required procedures.[121] Rather, the appropriate process should consider and select a set of potentially useful procedures based on the characteristics of the particular matter, such as the severity of the deprivation and the government interest at stake.[122] He also notes that civil procedural due process has moved beyond regulatory areas such as disability and welfare[123] For example, in 1959, in *Greene v. McElroy*, the Supreme Court held that the government had violated an organization's due process rights when the government identified it as Communist and subversive without an opportunity to be heard.[124]

The recognition of these "stigmatic" liberty interests was a key turning point in the expansion of civil procedural due process in the American courts.[125] Moreover, it has profound implications for data due process because predictive privacy harms often have the potential for stigmatic results. For instance, recent practices in commercial aviation illustrate similar violations of due process as those in *Greene*.[126] The similarities between the

---

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

[121] *Id.* at 1268–70.

[122] *Id.* at 1269–70. In describing the flexibility of the common law "fair procedure" requirement, Judge Friendly provided:

> The common law requirement of a fair procedure does not compel formal proceedings with all the embellishments of a court trial . . . nor adherence to a single mode of process. It may be satisfied by any one of a variety of procedures which afford a fair opportunity for an applicant to present his position. As such, this court should not attempt to fix a rigid procedure that must invariably be observed. Instead, the associations themselves should retain the initial and primary responsibility for devising method which provides an applicant adequate notice of the "charges" against him and a reasonable opportunity to respond.

Id. at 1270 n.10 (quoting Pinsker v. Pac. Coast Soc'y of Orthodontists, 526 P.2d 253, 263–64 (Cal. 1974) (en banc)).

[123] *Id.* at 1273. Judge Friendly also pointed to cases involving public education. *Id.* at 1274–75 (citing Wood v. Strickland, 420 U.S. 308 (1975); Perry v. Sindermann, 408 U.S. 593 (1972); Bd. of Regents of State Colleges v. Roth, 408 U.S. 564 (1972)). For instance, in 1972, in *Perry v. Sindermann*, the Supreme Court held that a public institution could not dismiss a professor without a hearing if he had tenure or if the dismissal would impair his ability to obtain future employment. 408 U.S. at 596.

[124] 360 U.S. 474, 507–08 (1959); Friendly, *supra* note 120, at 1273.

[125] *See* Friendly, *supra* note 120, at 1274–75.

[126] *Compare* 360 U.S. at 507–08 (holding that governmental labeling of an organization as Communist and subversive without an opportunity to be heard was a due process violation), *with* Danielle K. Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1256 (2008) (noting

harms addressed by due process in *Greene* and the errors and mistaken assumptions that have been revealed about the Transportation Security Administration's "No Fly" list—another product of Big Data—further support the need for data due process.[127]

On the other hand, as Judge Friendly notes, it is important to recognize a balance between protecting individual rights through due process and providing the costly administrative measures to do so.[128] Accordingly, similar onerous costs should not be imposed on Big Data providers because they would slow down the process of administering justice as well as encourage harassment and game-playing. This balance of protection with cost will also itself be dynamic and thus better considered as a standard than a rule—yet another reason why a due process approach is well-suited for Big Data.[129]

So, what kind of hearing is required for data due process? Judge Friendly writes that the affected party must receive an opportunity to present an argument, evidence, and corrections to prejudice.[130] In this regard, the required level of procedural safeguards varies directly with the importance of the affected private interest, the need for that particular safeguard in the given circumstances, and its utility.[131] Furthermore, it varies inversely with the administrative burden and any other adverse consequenc-

---

that "[e]very week, approximately 1,500 airline travelers reportedly are mislabeled as terrorists due to errors in the data-matching program known as the 'No Fly' list").

[127] Citron, *supra* note 126, at 1256. Although some due process cases have held that reputational harms are more appropriate for the province of tort law, due process can apply when reputation leads to deprivation of liberty or property. *Compare* Paul v. Davis, 424 U.S. 693, 699–701 (1976) (denying due process claim over stigmatic harm related to future employment opportunities stemming from inclusion in a flyer of "active shoplifters"), *with* Wisconsin v. Constantineau, 400 U.S. 433, 436–37 (1971) (holding that a ban on distributing alcoholic drinks to persons whose names were "posted" as excessive drinkers was a deprivation of liberty because it altered or extinguished a distinct right previously recognized by state law).

[128] Friendly, *supra* note 120, at 1276 ("[A]t some point the benefit to individuals from an additional [procedural] safeguard is substantially out-weighed by the cost of providing such protection, and that the expense of protecting those likely to be found undeserving will probably come out of the pockets of the deserving.").

[129] *See* Citron, *supra* note 126, at 1301–03 (discussing the tension between standards and rules).

[130] *See* Friendly, *supra* note 120, at 1277 ("A hearing in its very essence demands that he who is entitled to it shall have the right to support his allegations by argument however brief, and, if need be, by proof, however informal." (quoting Londoner v. Denver, 210 U.S. 373, 386 (1908))). To support his view, Judge Friendly referenced such authority as Justice Felix Frankfurter's concurring opinion in *McGrath*, that "even in the case of 'a person in jeopardy of serious loss,' that one must be given 'notice of the case against him and opportunity to meet it.'" *Id.* (quoting 341 U.S. at 171–72 (Frankfurter, J., concurring)). Judge Friendly further relied upon English common law, which characterized due process as "a fair opportunity . . . for correcting or contradicting any relevant statement prejudicial to [one's] view." *Id.* (quoting Board of Educ. v. Rice, [1911] A.C. 179, 182).

[131] *Id.* at 1278.

es.[132] To offset this balancing test's uncertainty, Judge Friendly suggests that "more elaborate specification of the relevant factors may help to produce more principled and predictable decisions."[133] It is this sense of principle and predictability that inspires this Article to bring due process to Big Data and its potential privacy harms.

Having laid out his general vision of due process hearings, Judge Friendly then goes on to enunciate eleven potential elements of a hearing that may help ensure a fair process. Not all are required, he states, but all are worth consideration depending on the circumstances at issue. They are: (1) an unbiased tribunal;[134] (2) notice of the proposed action and the grounds asserted for it;[135] (3) an opportunity to present reasons why the proposed action should not be taken;[136] (4) the right to call witnesses;[137] (5) the right to know the evidence against oneself;[138] (6) the right to have the decision based only on the evidence presented;[139] (7) the right to counsel;[140] (8) the making of a record;[141] (9) a statement of reasons;[142] (10) public attendance;[143] and (11) judicial review.[144]

Not all of these eleven elements would fit a data due process, of course. The right to call witnesses, for example, would be difficult and potentially cumbersome given how Big Data systems perform their analyt-

---

[132] *Id.*

[133] *Id.*

[134] *Id.* at 1279 ("Although an unbiased tribunal is a necessary element in every case where a hearing is required, sharp disagreement can arise over how much in the way of prior participation constitutes bias.").

[135] *Id.* at 1280–81 ("It is likewise fundamental that notice be given and that it be timely and clearly inform the individual of the proposed action and the grounds for it.").

[136] *Id.* at 1281.

[137] *Id.* at 1282 ("Under most conditions there does not seem to be any disposition to deny the right to call witnesses, although the tribunal must be entitled reasonably to limit their number and the scope of examination.").

[138] *Id.* at 1283–87 (noting disagreement as to whether "the right to know the nature of the evidence on which the administrator relies" applies only to criminal cases or to cases including administrative and regulatory actions).

[139] *Id.* at 1284–87 (discussing the necessity to grant an opportunity to confront all evidence and witnesses).

[140] *Id.* at 1287–91 (emphasizing the importance of counsel's role "to advance his client's cause by any ethical means").

[141] *Id.* at 1291–92 (highlighting the importance of a record but cautioning against "the sheer problem of warehousing these mountains of papers").

[142] *Id.* (finding a written statement of reasons necessary for purposes of judicial review; to provide for justification as a powerful preventive of wrong decisions; to encourage uniformity across decision-making bodies; and to make decisions somewhat more acceptable to a losing claimant).

[143] *Id.* at 1293–94 (citing three principal reasons for the right to an open trial as a part of due process: (1) fostering public confidence in the outcome; (2) helping to assure the accuracy of the evidence offered; and (3) placing pressure on the presiding officials to conduct the proceedings fairly). Judge Friendly, however, acknowledges that public attendance can also be disruptive in certain contexts such as prison disciplinary hearings. *Id.*

[144] *Id.* at 1294–95 (suggesting that judicial review be limited to questions of fair procedure).

ics.[145] On the other hand, elements such as "an unbiased tribunal," "the right to know the evidence against one," "the making of a record," and "a statement of reasons" make more sense for data due process. For example, rather than focusing—as FIPPs do—on the right to audit the personal data that has been collected about oneself generally, data due process would specifically focus on the right to audit the data used to make the determination at issue. Moreover, although both FIPPs and due process describe the concept of "notice" as critical, due process's notice focuses on the proposed action to be taken against the individual, rather than the type and amount of data to be collected or used—as is FIPPs focus. Again, because it is hard to predict in advance what processes or queries will be conducted by Big Data, due process's required notice of a proposed action fits Big Data's information practices better than FIPPs' approach to gatekeeping at the collection stage.[146]

An unbiased tribunal and judicial review would also be appropriate for data due process. Algorithmic bias is a serious issue, and there should be means for challenging it.[147] Because predictive privacy harms are often only discernable in hindsight, it may make sense to provide for some agency or judicial oversight when they occur. This oversight would only apply to data due process and not to the actual result, ensuring that the reviews would be fairly standardized and that the growing expertise of the agency or court performing these reviews would promote efficiency over the long-term.

2. The Nature of the Action

To address how the nature of the government action should influence the due process requirements, Judge Friendly argues that the greater the seriousness of the deprivation to the individual, the more protections should be in place.[148] For example, taking action against a citizen is far more serious than simply denying a citizen's request.[149] Among the deprivations he ranks as deserving the most procedural protection, Judge Friendly includes revocation of parole or probation, civil commitment, warrants, and revocation of a professional license.[150] He also suggests that gradations in deprivation matter:

---

[145] *See* Friendly, *supra* note 120, at 1282.

[146] *See* Citron, *supra* note 126, at 1305–06 (articulating that current notice is inadequate and suggesting improvements that mirror the notice requirements of other countries).

[147] *Id.* at 1262 (discussing possibility that programmers may distort policy with own biases when tasked with devising code to achieve policy's goals).

[148] *See* Friendly, *supra* note 120, at 1295–1304.

[149] *Id.*

[150] *Id.* at 1296–98.

> Thus a welfare termination is more serious than a reduction; suspension of a payment that is the claimant's only hope for income is more serious than a suspension that permits resort to other sources of income, even to the welfare system; expulsion from public housing is more serious than transfer to a smaller apartment; expulsion from a school is more serious than suspension or loss of credit; severance from government service is graver than suspension pending a further hearing; dismissal on a ground carrying a moral stigma is more serious than on one that does not; some types of discipline are more onerous than others.[151]

In terms of data due process, the type of predictive privacy harm should similarly influence the due process requirements. The greater the stigma or seriousness of the determination, the greater right one should have to question how Big Data adjudicated that result. For example, health information is among the most precious and protected, so more due process would be afforded to determinations in this field. Law enforcement uses would also be among those most subject to more scrutiny. Advertising might be on the lesser end of scrutiny. For mixed uses, such as the Target pregnancy example, which may be categorized as both advertising and health information, the greater protection should govern.[152]

C. *The Underlying Values of Due Process*

To further assess the appropriateness of a due process approach to Big Data, it is worth considering the values that underlie many due process rules. As Martin Redish and Larry Marshall have noted, due process's values remain more consistent than its procedures.[153] Application of due process to data requires significant imagination to design the appropriate processes and procedures. The large computational use of data will be varied and contextual (much like the range of cases that courts consider). Accordingly, a flexible model based more on values and less on specific procedures will be more likely to endure over time.[154]

---

[151] *Id.* at 1298 (footnotes omitted) (noting the impossibility of implementing a universal scale).

[152] *See* Duhigg, *supra* note 5, § 6 (Magazine), at 30 (discussing the Target example).

[153] Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L.J. 455, 474 (1986). Redish and Marshall stress that a procedural due process model should have flexible procedural mechanisms that maintain the due process clause's long-existing values. *Id.*

[154] *See id.* at 474–75.

In their examination of due process, Redish and Marshall set out seven enduring sets of values that due process should preserve: (1) accuracy;[155] (2) the appearance of fairness;[156] (3) equality of inputs into the process;[157] (4) predictability, transparency, and rationality;[158] (5) participation;[159] (6) revelation;[160] and (7) privacy-dignity.[161]

Each of these values maps well to our concerns about Big Data. For Big Data to deliver the answers we seek, it must be accurate and include all appropriate inputs equally to overcome any signal problems. Otherwise, Big Data may provide us with misleading conclusions.[162] Furthermore, before Big Data's role in decision making can gain greater social acceptance—especially within government—it must not only appear fair but also have an acceptable degree of predictability, transparency, and rationality. Without these values, we cannot trust Big Data to be part of governance.[163] Finally, participation, revelation, and privacy-dignity would help optimize Big Data's role in public decision making for the same reasons they optimize the judicial or administrative process by bringing legitimacy to the process, albeit through different approaches. These values address the individual's concern about the procedural process, even in spite of unfavorable outcomes.[164]

---

[155] *Id.* at 476–81 (explaining procedural mechanisms such as an independent adjudicator and right to counsel to ensure accuracy).

[156] *Id.* at 483–84. (characterizing an appearance of fairness via an independent adjudicator as the flip side of the accuracy value because it fosters trust in the adjudicatory process).

[157] *Id.* at 484–85 (emphasizing the importance of procedural fairness to ensure that a party's identity does not affect the adjudicatory process).

[158] *Id.* at 485–86 (reasoning that due process's predictability, transparency and rationality allow individuals to plan rationally and make informed decisions).

[159] *Id.* at 487–89 (attributing the value of participation to individual psychological benefits and societal benefits); *see* Marshall v. Jerrico, Inc., 446 U.S. 238, 242 (1980) (specifying "the promotion of participation and dialogue by affected individuals in the [decision-making] process" as a central concern of procedural due process).

[160] Redish & Marshall, *supra* note 153, at 489–91 (noting that, although this value is more introspective than others, it is pivotal to preserving individual dignity and understanding).

[161] *Id.* at 491 (suggesting that the Supreme Court acknowledges individual physical and mental privacy, but stating that this value restricts due process's procedural mechanisms).

[162] *See, e.g.*, Declan Butler, *When Google Got Flu Wrong*, 494 NATURE 155, 155–56 (2013) (explaining Google Flu Trends' faulty overestimation when it used Big Data to determine peak flu levels and the effects of mistaken reliance on its analysis); Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013, 2:00 PM), http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html, *archived at* http://perma.cc/B3U8-K67A (discussing Big Data's signal problems in such scenarios as analysis of GPS data to detect potholes).

[163] *See generally* Kate Crawford & Catherine Lumby, *Networks of Governance: Users, Platforms, and the Challenges of Networked Media Regulation*, 1 INT'L J. TECH. POL'Y & LAW 270 (2013). By the term "governance," we are referring primarily to networked or technological governance, which involves both governmental aspects as well as private and individual ones. *Id.*

[164] Robert J. MacCoun, *Voice, Control, and Belonging: The Double-Edged Sword of Procedural Justice*, 1 ANN. REV. L. & SOC. SCI. 171, 171–73 (2005) (highlighting "the ability to tell

Redish and Marshall also raise two cautionary concerns about what they consider to be the centerpiece of any due process framework—the independent adjudicator. Specifically, they highlight the dangers that arise when an adjudicator has either a direct financial interest in the proceeding's outcome or an inherent personal bias.[165] There can be no doubt that for-profit providers of Big Data analytics have direct financial interests in some of the outputs they produce. Furthermore, as we note in another article, issues of bias also exist within Big Data's algorithms and data sets, despite their appearance of objectivity.[166] These present two additional reasons to apply due process to these data regimes.

D. *Due Process as Separation of Powers and Systems Management*

Due process's historical role as a means of separating powers among governments is another favorable reason to consider it as a mechanism to address how Big Data handles personal information. Due process has ensured that those who pass general laws are kept separate from both those who are called upon to enforce them in specific circumstances, and those who judge whether or not those cases have merit. As two scholars write, this protects citizens against directed executive punishment in the form of adjudication.[167] Congress may pass laws affecting our lives, liberty, and property, and the President may sign them, but their enforcement requires a fair process overseen by a neutral arbiter.[168] Thus, a core function of due process is to separate those who write the legal code from adjudicators who use it.[169]

With many Big Data determinations, there is little or no regulation of the interactions among the algorithm's designer (the lawmaker), the person who oversees the queries (the executive), and the adjudicator (the computational output). Accordingly, there is no system of checks and balances to ensure that biases are not present in the system, which is especially crucial to a system of enforcement. As Chief Justice John Marshall has explained: "It is the peculiar province of the legislature to prescribe general rules for the government of society; the application of those rules to individuals in

---

one's story" and "dignified, respectful treatments" as significant dimensions of due process to individuals, regardless of the adjudicatory process's outcome).

[165] Redish & Marshall, *supra* note 153, at 494–505.

[166] *See* Crawford, *supra* note 162; Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 9, 2013), http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data, *archived at* http://perma.cc/67SQ-5BXK.

[167] Nathan S. Chapman & Michael W. McConnell, *Due Process as Separation of Power*, 121 YALE L.J. 1672, 1782–92 (2012).

[168] *Id.* at 1677–1726 (discussing the evolution of how due process acts as a separation of powers).

[169] *Id.*

society would seem to be the duty of other departments."[170] Due process would help ensure that Big Data does not blur its processes with its provinces.

Various due process scholars have also conceptualized the doctrine as a form of systematic management technique that should focus less on any individual harm and more on discovering errors, identifying their causes, and implementing corrective actions.[171] Or, as one scholar suggests, although due process should address injustices individually, it should look beyond them to the managerial level by creating schemes and incentives to normatively circumscribe government actions within the bounds of law.[172] Similarly, due process can serve as a systematic management technique for Big Data by uncovering errors, identifying their causes, and providing schemes and incentives to correct them while keeping within the bounds of privacy laws and norms.

III. TOWARD A MODEL FOR DATA DUE PROCESS

A. *Technological Due Process: The Citron Analysis*

The general idea of applying due process to automated systems is not new.[173] In her 2010 article, *Technological Due Process*, Danielle Citron examines the use of automated systems in governmental administrative proceedings, the risks they pose to deprivations of liberty and property, and how a reinvigorated approach to due process could help mitigate and ad-

---

[170] Fletcher v. Peck, 10 U.S. (6 Cranch) 87, 136 (1810); *see* Chapman & McConnell, *supra* note 167, at 1733 (emphasizing the importance of applying the laws equally to all people; suggesting that separation of powers serves as an effective vehicle to accomplish this; and using Chief Justice Marshall's opinion to support this argument).

[171] *See, e.g.*, Citron, *supra* note 126, at 1301–13; Richard H. Fallon, Jr., *Some Confusion About Due Process, Judicial Review, and Constitutional Remedies*, 93 COLUM. L. REV. 309, 311, 336–37 (1993); Jerry L. Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772, 815–16 (1974).

[172] Fallon, *supra* note 171, at 311.

[173] *See, e.g.*, Citron, *supra* note 126, 1301–13; Ian Kerr, *Prediction, Pre-emption, Presumption: 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, 107 (Mireille Hildebrandt & Katja de Vries eds., 2013) (noting that "from a broad legal and ethical perspective, problems are sure to arise when anticipatory algorithms and other computational systems import norms that undermine the due process otherwise afforded to citizens by law"); Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 64–81 (2005) (addressing the due process effects of using data matching and mining to identify persons against whom official action is taken, such as in the use of air passenger screening and the maintenance of watch lists); Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1553–68 (critiquing traditional notions of transparency in the context of predictive analytics and advocating for strong procedural protections to counteract transparency's inability to effectively hold government actors accountable for their use of Big Data).

dress these problems.[174] Citron's approach could be expanded to address the predictive privacy harms of Big Data.

First, Citron identifies various automated systems that government administrative officials use to adjudicate individual liberty or property interests. [175] These include systems that terminate Medicaid, food stamps, and other welfare benefits; target people for exclusion from air travel; identify parents who neglect child support payments; purge voters from rolls without notice; and deem small businesses ineligible for federal contracts.[176] She also notes that most of these systems (1) have failed to give adequate notice to individuals whose interests were at stake; (2) have failed to provide any opportunity to be heard before a decision was rendered; and (3) have often adjudicated the case in secrecy or without leaving any record for audits or judicial review.[177]

In particular, Citron notes that automatic systems generally fail to give any or adequate notice to individuals when their liberty or property interests are algorithmically adjudicated.[178] In administrative proceedings, notice of an action against one's liberty interest should "be 'reasonably calculated' to inform . . . affected individuals of the issues to be decided, the evidence supporting the government's position, and the agency's decisional process."[179] When affected individuals do not receive this adequate notice, they lack sufficient information to respond effectively to the claim.[180] As Citron writes, clear notice should decrease the likelihood that agency action will rely on false premises, misleading presumptions, or misapplication of rules.[181] To counteract this failure to give notice, Citron argues that automated administrative systems must include audit trails that record the facts and rules supporting each decision.[182] This trail can then be compiled into some form of sufficient notice when a decision is made, and subsequently transmitted to the affected individual.[183]

Big Data systems suffer from many of the same weaknesses as government administration systems regarding notice. Individuals or groups that are subjected to predictive privacy harms rarely receive any meaningful notice of the predictions before they occur or are implemented; and even then,

---

[174] Citron, *supra* note 126, 1251–58.
[175] *Id.* at 1252.
[176] *Id.*
[177] *Id.* at 1279–83.
[178] *Id.* at 1281.
[179] *Id.* at 1281–82.
[180] *Id.* at 1282 (citing Cosby v. Ward, 843 F.2d 967, 984 (7th Cir. 1988)).
[181] *Id.* at 1282 (citing Goldberg v. Kelly, 397 U.S. 254, 268 (1970)); *see also* Tene & Polonetsky, *supra* note 14, at 271 (emphasizing the need for full disclosure of the process and criteria used when making decisions that affect individuals' lives).
[182] Citron, *supra* note 126, at 1305.
[183] *Id.*

providers are unlikely to share the evidence and reasoning for the predictions that were made. Notably, there is currently no legal requirement that providers archive any audit trail or retain any record of the basis of the prediction.

Opportunities to be heard also present problems for automated systems and due process.[184] Citron posits that an "opportunity to be heard" in a Big Data context would involve access to an automated program's source code or a hearing on the logic of a computer program's decision, and it would often be found far too expensive under the *Mathews* balancing test.[185] She notes, however, that because such challenges would encourage fairness in future cases, it might be worth pursuing an opportunity to be heard even under *Mathews*.[186]

To better adapt due process application to automated systems, Citron proposes several changes. She suggests first that instead of subjecting every automated system to cross-examination, one could—at a minimum— invest in extra education about the biases and fallacies of automation for government personnel who use the systems to make administrative decisions.[187] Educating these individuals about the systems' flaws could help them scrutinize the outputs more fairly.[188] Second, she suggests that agencies should require hearing officers "to explain, in detail, their reliance on an automated system's decision," including any computer-generated facts or legal findings.[189] Third, she suggests that agencies should be required to regularly test their system's software for bias and other errors.[190]

For Big Data adjudications, many of these same problems exist with algorithmic biases and the potential to inaccurately predict PII about individuals. Thus, similar opportunities to be heard may well be appropriate, especially with respect to educating data scientists about the biases of Big Data, requiring those who use Big Data for significant decisions concerning individuals to disclose which data sets were used, and requiring testing of predictive analytics to assess how accurate a given system can be.

---

[184] *Id.* at 1283 (noting that the *Mathews v. Eldridge* balancing test aspires to provide an opportunity to be heard "at a meaningful time and in a meaningful manner," but only if it is cost-efficient (citing 424 U.S. 319, 333 (1976))).

[185] *Id.* at 1284.

[186] *Id.* at 1285–87. For example, Citron acknowledges that certain automated systems, such as the "No Fly" list, involve state secrets and would rarely be subjected to scrutiny. *Id.* at 1286.

[187] *Id.* at 1306.

[188] *Id.* at 1306 (noting the success of special scientific theory and methodology workshops for federal district court judges who need to assess the reliability of expert testimony).

[189] *Id.* at 1307.

[190] *Id.* at 1310.

Finally, Citron discusses what meaningful judicial review for automated systems might entail and why most of these systems evade it.[191] Specifically, she critiques automated administrative systems because they often fail to retain any audit record of how they made the decisions at issue or upon what data the decision was based.[192] Again, similar to the need for notice, Big Data may benefit from an audit trail because it provides reassurance and increases accuracy. Access to audit trails would also allow individuals to raise specific objections to how and when their data is being used in various processes.

B. *Procedural Data Due Process*

As noted above, procedural due process generally describes the constitutional requirement that any government deprivation of a liberty or property right must be preceded—at a minimum—by notice and the opportunity for a hearing on the matter before an impartial adjudicator. In thinking about procedural data due process, this Section will draw from these same three elements, while incorporating aspects of Judge Friendly's list of eleven, and Redish and Marshall's values of due process.[193]

To begin, some uses of Big Data will be difficult to fit in the mold of individualized due process adjudication. These uses include the opportunities individuals were not selected for, the advantageous insurance advertising offers that did not appear in their search results, and the jobs they never knew existed because they didn't fit the desired profile of a marketer. But when individuals are aware of or directly involved in processes in which Big Data is used as part of the outcome, such as when it is used to identify top candidates from a given pool of applicants, individualized due process approaches will seem most appropriate. For the more opaque predictive problems—including missed opportunities, such as a real estate offer one never sees because Big Data might have judged one unworthy—a more structural due process approach might be better, with oversight and auditing primarily driven by public agencies. An alternative could be granted in the form of a remedial tort, with standing and statutory damages for those

---

[191] *Id.* at 1276–77 (noting that designers of automated systems have intentionally chosen not to include audit trails in their systems); *id. at* 1298–1300 (explaining that audit trails for automated systems would help officials understand system failures when they occur).

[192] *Id.* at 1300.

[193] *See* Friendly, *supra* note 120; Redish & Marshall, *supra* note 153; *see also* Chapman & McConnell, *supra* note 167, at 1774 (highlighting Supreme Court's emphasis that "to comply with due process, statutes must either provide for the use of common law procedures or, if they do not, employ alternative procedures that the courts would regard as equivalently fair and appropriate").

whose rights have been violated, like the scheme set forth in the ECPA.[194] This could also include class certification and association standing options.

Another key question arises as to when due process should attach to a particular decision. Although the exact moment for any particular decision is too specific for this Article to discuss, one can imagine that certain determinations, like determinations of employment eligibility, will be more regularized and repetitive.[195] Thus, due process could attach at the moment the decision is made to determine the specific eligibility of a particular set of individuals. The moment of attachment could also be triggered sooner as the generated data approaches the equivalent of PII. The more closely Big Data resembles PII-type information, the stronger the case for attaching procedural data due process. In addition, the greater the seriousness of the decision, the more Big Data due process is afforded. With these questions in mind, the remainder of this Section turns to principles for implementation.

1. Notice

Our conception of notice for procedural data due process centers on providing those who may suffer from predictive privacy harms an opportunity to intervene in the predictive process. This opportunity ensures fairness with respect to the processes that affect their interests, either individually or structurally. Our approach would require those who use Big Data to "adjudicate" others—i.e., those who make categorical or attributive determinations—to post some form of notice, disclosing not only the type of predictions they attempt, but also the general sources of data that they draw upon as inputs, including a means whereby those whose personal data is included can learn of that fact.

One could also imagine a variety of notice rights and obligations that would enable consumers to petition Big Data providers to check and see if their data was being included or used in any predictive adjudications, and whether that data was accurate. Similar laws are available for data collection.[196] These models could be expanded to include data processing for Big

---

[194] *See* 18 U.S.C. §§ 2510–2522 (2012).

[195] *See, e.g.*, Eric Markowitz, *Meet a Start-Up With a Big Data Approach to Hiring*, INC., http://www.inc.com/eric-markowitz/how-data-can-help-you-recruit-talented-engineers.html, *archived at* http://perma.cc/4HVV-39TJ (last updated Sept. 19, 2013) (describing the development of a "tech talent-sourcing platform" using Big Data); Joseph Walker, *Meet the New Boss: Big Data*, WALL ST. J. (Sept. 20, 2012, 11:16AM), http://online.wsj.com/article/SB10000872396390443890304578006252019616768.html, *archived at* http://perma.cc/SR7X-UNXT (discussing companies' use of Big Data for hiring decisions).

[196] *See, e.g.*, Online Privacy Protection Act of 2003, CAL. BUS. & PROF. CODE §§ 22575–22579 (West 2008) (amended 2013) (requiring commercial operators of online services to conspicuously post privacy policies informing users of what PII is collected and how it will be used);

Data so that, like privacy policies disclosing data collection practices, they would disclose data prediction practices "reasonably calculated" to inform individuals of the risks to which they may be exposed in terms of predictive privacy harms. Moreover, when a particular set of predictions about an individual or discrete group has been queried (or "adjudicated"), notice would be sent out that is "reasonably calculated" to inform those affected of, at a minimum, the issues that were predicted, and ideally, the data considered and the methodology employed. At a minimum, this notice should provide for a mechanism to access the audit trail or record created in the predictive process.[197]

For example, if a company were to license search query data from Google and Bing in order to predict which job applicants would be best suited for a particular position, it would have to disclose to all applicants that it uses search queries for predictive analytics related to their candidacy. Or in the case of predictive policing, the government would have to notify citizens that it was using predictive analytics and particular sets of public records to determine which areas of a city it marked as "hot spots" as well as its capacity to determine if one lived or worked within the actual hot spots.

Another example would focus on the issue of fair housing. If landlords and real estate companies were to shift away from general advertising in media outlets and toward using Big Data to determine likely buyers or renters who fit their "ideal" profiles, we could, again, require them to disclose this practice. Depending on the specifics of the practice, one could imagine the notice either on an individual level—to those who knew of their inclusion in the predictions—or on a structural level, if the predictions were for a large set of a given population.

2. Opportunity for a Hearing

Once notice is available, the question then becomes how one might challenge the fairness of the predictive process employed. We believe that the most robust mechanism for this is the opportunity to be heard and, if

---

MINN. STAT. §§ 325M.01–.09 (2012) (prohibiting the disclosure of PII and requiring online service providers to get users' permission before disclosing their internet history).

[197] *See* Citron, *supra* note 126, at 1305–06 (discussing the need for audit trails for the government's administrative technological systems to facilitate meaningful notice to individuals); Dwork & Mulligan, *supra* note 52, at 38–39 (suggesting that bias testing and transparency of Big Data analytics may mitigate privacy harms); *see also* Consultative Comm. of the Convention for the Prot. of Individuals with Regard to Automatic Processing of Pers. Data [ETS No. 108], *Propositions of Modernisation*, COUNCIL OF EUR. 4–5 (Dec. 18, 2012), http://www.coe.int/t/dghl/standardsetting/dataprotection/TPD_documents/T-PD(2012)04Rev4_E_Convention%20108%20modernised%20version.pdf, *archived at* http://perma.cc/N2QZ-QV8F (discussing "Rights of the data subject").

necessary, correct the record. This would include examining the evidence used, including both the data input and the algorithmic logic applied. In contexts in which security and proprietary concerns arise—or in more structural situations—this role could be given to a trusted third party who would act as a neutral data arbiter to routinely examine Big Data providers whose adjudications give rise to predictive privacy harms. For example, the FTC—which currently addresses many privacy harms involving technology and has recently hired technologists to assist its investigations and enforcement actions—could investigate complaints based on predictive privacy harms and, in the process of those complaints, investigate the basis of the predictions.[198]

The presence of a neutral data arbiter would provide the public with an opportunity to be heard, to examine the evidence used in adjudicative predictions, and to challenge it. This approach would also comport with several of the underlying values of due process: accuracy of the determination; appearance of fairness; predictability, transparency, and rationality; participation; and revelation. In particular, because Big Data generally excludes any user participation in its decision making, a neutral data arbiter would be especially important to ensure that there was a meaningful hearing for public concerns.

3. Impartial Adjudicator and Judicial Review

One of the primary myths about Big Data is that it produces outputs that are somehow free from bias and closer to objective truth than other forms of knowledge.[199] Due process requires that those who deprive individuals of a liberty interest do so without unwarranted bias or direct financial interest in the outcome. Procedural data due process, therefore, can also serve as a valuable framework for ensuring greater fairness with predictive analytics. A neutral data arbiter could field complaints and investigate sufficient allegations of bias or financial interest that might render the adjudication unfair. In particular, drawing on the literature exploring due process as a function of separation of powers, the arbiter could examine the relationship between those who designed the analytics and those who run the individual processes to make sure that their roles are appropriate and distinct. This would require some form of audit trail that records the basis of predic-

[198] *See* Press Release, Fed. Trade Comm'n, FTC Names Edward W. Felten as Agency's Chief Technologist; Eileen Harrington as Executive Director (Nov. 4, 2010), http://www.ftc.gov/opa/2010/11/cted.shtm, *archived at* http://perma.cc/EEY2-TJHV. This would also address concerns about standing, in which a single plaintiff might not have sufficient evidence to show an individual concrete harm in a particular case without gathering evidence through the litigation discovery process.

[199] boyd & Crawford, *supra* note 15, at 667.

tive decisions, both in terms of the data used and the algorithm employed. Such audits are already used in various data-mining contexts and, thus, would not be unreasonable to require.[200]

CONCLUSION

In concluding his article on hearings, Judge Friendly wrote: "We have traveled over wide areas—from termination of welfare payments to the establishment of incentive per diem for freight cars, from student and prison discipline to rates for natural gas. Yet the problem is always the same—to devise procedures that are both fair and feasible."[201] This Article ends here with the same observation. Big Data presents many challenges for privacy, to which this Article posits a model of procedural data due process as a response. How exactly it responds to each challenge may vary, but it will ultimately succeed if it can ensure protections that are both fair and feasible for those at risk from this new form of privacy harm.

---

[200] *See* Gray & Citron, *supra* note 64, at 118–19 (noting that the predictive analytic systems of New York City's DAS and Palantir both provide mechanisms for outside review). On the question of remedies, with procedural safeguards, one can imagine several common remedies for curing a violation. In the judicial system, a specific proceeding or determination might be invalidated, thereby forcing the agency or adjudicator to revisit the determination using proper processes. In the privacy context, there is some precedent for this in France. *See* Loi 2004-801 du 6 août 2004 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel et modifiant la loi 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (1) [Law 2004-801 of August 6, 2004 regarding the Protection of Individuals Regarding their Personal Data and modifying Law 78-17 relating to Data Processing, Files, and Freedoms], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Aug. 6, 2004, p. 14063.

[201] Friendly, *supra* note 120, at 1315.