66 STAN. L. REV. 25
September 3, 2013

PRIVACY AND BIG DATA: MAKING ENDS MEET

Jules Polonetsky & Omer Tene*

INTRODUCTION

How should privacy risks be weighed against big data rewards? The recent controversy over leaked documents revealing the massive scope of data collection, analysis, and use by the NSA and possibly other national security organizations has hurled to the forefront of public attention the delicate balance between privacy risks and big data opportunities.[1] The NSA revelations crystalized privacy advocates' concerns of "sleepwalking into a surveillance society" even as decisionmakers remain loath to curb government powers for fear of terrorist or cybersecurity attacks.

Big data creates tremendous opportunity for the world economy not only in the field of national security, but also in areas ranging from marketing and credit risk analysis to medical research and urban planning. At the same time, the extraordinary benefits of big data are tempered by concerns over privacy and data protection. Privacy advocates are concerned that the advances of the data ecosystem will upend the power relationships between government, business, and individuals, and lead to racial or other profiling, discrimination, overcriminalization, and other restricted freedoms.

---

* Jules Polonetsky is Co-Chair and Director, Future of Privacy Forum. Omer Tene is Associate Professor, College of Management Haim Striks School of Law, Israel; Senior Fellow, Future of Privacy Forum; Affiliate Scholar, Stanford Center for Internet and Society. We would like to thank Joseph Jerome, Legal and Policy Fellow at the Future of Privacy Forum, for his research assistance.

1. Glenn Greenwald, *NSA Collecting Phone Records of Millions of Verizon Customers Daily*, GUARDIAN (June 6, 2013), http://www.guardian.co.uk/world/2013/jun/06/nsa-phone-records-verizon-court-order; Glenn Greenwald & Ewen MacAskill, *NSA Prism Program Taps in to User Data of Apple, Google and Others*, GUARDIAN (June 7, 2013), http://www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data.

Finding the right balance between privacy risks and big data rewards may very well be the biggest public policy challenge of our time.[2] It calls for momentous choices to be made between weighty policy concerns such as scientific research, public health, national security, law enforcement, and efficient use of resources, on the one hand, and individuals' rights to privacy, fairness, equality, and freedom of speech, on the other hand. It requires deciding whether efforts to cure fatal disease or eviscerate terrorism are worth subjecting human individuality to omniscient surveillance and algorithmic decision-making.[3]

Unfortunately, the discussion progresses crisis by crisis, often focusing on legalistic formalities while the bigger policy choices are avoided. Moreover, the debate has become increasingly polarized, with each cohort fully discounting the concerns of the other. For example, in the context of government surveillance, civil libertarians depict the government as pursuing absolute power, while law enforcement officials blame privacy for child pornography and airplanes falling out of the sky. It seems that for privacy hawks, no benefit no matter how compelling is large enough to offset privacy costs, while for data enthusiasts, privacy risks are no more than an afterthought in the pursuit of complete information.

This Essay suggests that while the current privacy debate methodologically explores the *risks* presented by big data, it fails to untangle commensurate *benefits*, treating them as a hodgepodge of individual, business, and government interests. Detailed frameworks have developed to help decisionmakers understand and quantify privacy risk, with privacy impact assessments now increasingly common for government and business undertakings.[4] Yet accounting for *costs* is only part of a balanced value equation. In order to complete a cost-benefit analysis, privacy professionals need to have at their disposal tools to assess, prioritize, and to the extent possible, quantify a project's *rewards*. To be sure, in recent years there have been thorough expositions of big data benefits.[5]

---

2. Ira Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 77-78 (2013); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 240-42 (2013).

3. We are not arguing that these public policy objectives are mutually exclusive. To the contrary, we support the "Privacy by Design" paradigm that aims to integrate privacy safeguards into projects, products, and services. Yet at some point, stark policy choices need to be made—this is where privacy costs need to be balanced against big data benefits. *See* Ann Cavoukian, *Privacy by Design: The Seven Foundational Principles*, INFO. PRIVACY COMM'R, ONT., CAN. (Jan. 2011), http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf ("*Privacy by Design* seeks to accommodate all legitimate interests and objectives in a positive-sum 'win-win' manner, not through a dated, zero-sum approach, where unnecessary trade-offs are made.").

4. *See, e.g.*, PRIVACY IMPACT ASSESSMENT 4-5 (David Wright & Paul De Hert eds., 2012); *Privacy Impact Assessments: The Privacy Office Official Guidance*, DEP'T HOMELAND SEC. (June 2010), http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_guidance_june2010.pdf.

5. *See, e.g.*, VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2013); RICK

But the societal value of these benefits may depend on their nature, on whether they are certain or speculative, and on whether they flow to individuals, communities, businesses, or society at large.

The integration of benefit considerations into privacy analysis is not without basis in current law. In fact, it fits neatly within existing privacy doctrine under both the FTC's authority to prohibit "unfair trade practices" in the United States[6] as well as the "legitimate interests of the controller" clause in the European Union data protection directive.[7] Over the past few years, the FTC has carefully recalibrated its section 5 powers to focus on "unfair" as opposed to "deceptive" trade practices. An "unfair" trade practice is one that "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and *is not outweighed by countervailing benefits* to consumers or competition."[8] Clearly, benefit considerations fit squarely within the legal analysis. Moreover, in determining whether an injury is outweighed by countervailing benefits, the FTC typically considers not only the impact on specific consumers but also on society at large.[9]

In the European Union, organizations are authorized to process personal data without an individual's consent based on such organizations' "legitimate interests" as balanced against individuals' privacy rights. In such cases, individuals have a right to object to processing based "on compelling legitimate grounds."[10] Similar to the FTC's "unfairness" doctrine, legitimate interest analysis is inexorably linked to a benefit assessment.

This Essay proposes parameters for a newly conceptualized cost-benefit equation that incorporates both the sizable benefits of big data as well as its attendant costs. Specifically, it suggests focusing on *who* are the beneficiaries of big data analysis, *what* is the nature of the perceived benefits, and with what level of *certainty* can those benefits be realized. In doing so, it offers ways to take account of benefits that accrue not only to businesses but also to individuals and to society at large.

SMOLAN & JENNIFER ERWITT, THE HUMAN FACE OF BIG DATA (2012); Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63 (2012); *Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance*, CTR. FOR INFO. POL'Y LEADERSHIP (Feb. 2013), http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf; *Unlocking the Value of Personal Data: From Collection to Usage*, WORLD ECON. F. (Feb. 2013), http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf.

6. 15 U.S.C. § 45(a)(1) (2011).

7. Council Directive 95/46, art. 7(f), 1995 O.J. (L 281) 31, 40 (EC), *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF.

8. 15 U.S.C. § 45(n) (emphasis added).

9. Woodrow Hartzog & Daniel Solove, The FTC and the New Common Law of Privacy (Aug. 19, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312913.

10. Council Directive, *supra* note 7, at art. 14(a).

I. BENEFICIARIES

Who benefits from big data? In examining the value of big data, we start by evaluating who is affected by the relevant breakthrough. In some cases, the individual whose data is processed directly receives a benefit. In other cases, the benefit to the individual is indirect. And in many other cases, the relevant individual receives no attributable benefit, with big data value reaped by business, government, or society at large.

A. *Individuals*

In certain cases, big data analysis provides a direct benefit to those individuals whose information is being used. This provides strong impetus for organizations to argue the merits of their use based on their returning value to affected individuals. In a previous article, we argued that in many such cases, relying on individuals' choices to legitimize data use rings hollow given well-documented biases in their decisionmaking processes.[11] In some cases, a particular practice may be difficult to explain within the brief opportunity that an individual pays attention, while in others, individuals may decline despite their best interests. Yet it would be unfortunate if failure to obtain meaningful consent would automatically discredit an information practice that directly benefits individuals.

Consider the high degree of customization pursued by Netflix and Amazon, which recommend films and products to consumers based on analysis of their previous interactions. Such data analysis directly benefits consumers and has been justified even without solicitation of explicit consent. Similarly, Comcast's decision in 2010 to proactively monitor its customers' computers to detect malware,[12] and more recent decisions by Internet service providers including Comcast, AT&T, and Verizon to reach out to consumers to report potential malware infections, were intended to directly benefit consumers.[13] Google's autocomplete and translate functions are based on comprehensive data collection and real time keystroke-by-keystroke analysis. The value proposition to consumers is clear and compelling.

In contrast, just *arguing* that data use benefits consumers will not carry the day. Consider the challenges that proponents of behavioral advertising have faced in persuading regulators that personalized ads deliver direct benefits to

---

11. Omer Tene & Jules Polonetsky, *To Track or 'Do Not Track': Advancing Transparency and Individual Control in Online Behavioral Advertising*, 13 MINN. J.L. SCI. & TECH. 281, 285-86 (2012).

12. Roy Furchgott, *Comcast to Protect Customer's Computers from Malware*, N.Y. TIMES GADGETWISE (Sept. 30, 2010), http://gadgetwise.blogs.nytimes.com/2010/09/30/comcast-to-monitor-customer-computers-for-malware.

13. Daniel Lippman & Julian Barnes, *Malware Threat to Internet Corralled*, WALL ST. J. (July 9, 2012), http://online.wsj.com/article/SB10001424052702303292204577515262710139518.html.

individuals. Behavioral ads are served by grouping audiences with specific web surfing histories or data attributes into categories, which are then sold to advertisers using algorithms designed to maximize revenue. Consumers may or may not perceive the resulting ads as relevant, and even if they do, they may not appreciate the benefit of being targeted with relevant ads.

B. *Community*

In certain cases, the collection and use of an individual's data benefits not only that individual, but also members of a proximate class, such as users of a similar product or residents of a geographical area. Consider Internet browser crash reports, which very few users opt into not so much because of real privacy concerns but rather due to a (misplaced) belief that others will do the job for them. Those users who do agree to send crash reports benefit not only themselves, but also other users of the same product. Similarly, individuals who report drug side effects confer a benefit to other existing and prospective users.[14]

C. *Organizations*

Big data analysis often benefits those organizations that collect and harness the data. Data-driven profits may be viewed as enhancing allocative efficiency by facilitating the "free" economy.[15] The emergence, expansion, and widespread use of innovative products and services at decreasing marginal costs have revolutionized global economies and societal structures, facilitating access to technology and knowledge[16] and fomenting social change.[17] With more data, businesses can optimize distribution methods, efficiently allocate credit, and robustly combat fraud, benefitting consumers as a whole.[18] But in the absence of individual value or broader societal gain, others may consider enhanced business profits to be a mere value transfer from individuals whose data is being exploited. In economic terms, such profits create distributional gains to some actors (and may in fact be socially regressive) as opposed to driving allocative efficiency.

---

14. Nicholas P. Tatonetti et al., *A Novel Signal Detection Algorithm for Identifying Hidden Drug-Drug Interactions in Adverse Event Reports*, 19 J. AM. MED. INFORMATICS ASS'N 79, 79-80 (2012).

15. CHRIS ANDERSON, FREE: THE FUTURE OF A RADICAL PRICE (2009).

16. Tim Worstall, *More People Have Mobile Phones than Toilets*, FORBES (Mar. 23, 2013), http://www.forbes.com/sites/timworstall/2013/03/23/more-people-have-mobile-phones-than-toilets.

17. WAEL GHONIM, REVOLUTION 2.0: THE POWER OF THE PEOPLE IS GREATER THAN THE PEOPLE IN POWER: A MEMOIR (2012).

18. *A Different Game: Information Is Transforming Traditional Businesses*, ECONOMIST (Feb. 25, 2010), http://www.economist.com/node/15557465.

D. *Society*

Finally, some data uses benefit society at large. These include, for example, data mining for purposes of national security. We do not claim that such practices are always justified; rather, that when weighing the benefits of national security driven policies, the effects should be assessed at a broad societal level. Similarly, data usage for fraud detection in the payment card industry helps facilitate safe, secure, and frictionless transactions, benefiting society as a whole. And large-scale analysis of geo-location data has been used for urban planning, disaster recovery, and optimization of energy consumption.

E. *Benefits*

Big data creates enormous value for the global economy, driving innovation, productivity, efficiency, and growth. Data has become the driving force behind almost every interaction between individuals, businesses, and governments. The uses of big data can be transformative and are sometimes difficult to anticipate at the time of initial collection. And any benefit analysis would be highly culture-specific. For example, environmental protection may be considered a matter of vital importance in the United States, but less so in China.

In a recent article titled *The Underwhelming Benefits of Big Data*, Paul Ohm critiques our previous articles, arguing that "Big Data's touted benefits are often less significant than claimed and less necessary than assumed."[19] He states that while some benefits, such as medical research, are compelling, others yield only "minimally interesting results."[20] He adds, "Tene and Polonetsky seem to understand the speciousness of some of the other benefits they herald."[21]

While we agree that society must come up with criteria to evaluate the relative weight of different benefits (or social values), we claim that such decisions transcend privacy law. The social value of energy conservation, law enforcement, or economic efficiency is a meta-privacy issue that requires debate by experts in the respective fields. If privacy regulators were the sole decision-makers determining the relative importance of values that sometimes conflict with privacy, such as free speech, environmental protection, public health, or national security, they would become the *de facto* regulators of all things commerce, research, security, and speech.[22] This would be a perverse result,

19. Paul Ohm, *The Underwhelming Benefits of Big Data*, 161 U. PA. L. REV. ONLINE 339, 340 (2013).

20. *Id*. at 344.

21. *Id*.

22. Currently, privacy regulators appear to be making almost arbitrary decisions when it comes to balancing privacy risks against potential data rewards. In fact, the recent Opinion of the Article 29 Working Party, which required national regulators to assess compatibility "on a case-by-case basis[,]" appears to legitimize an unpredictable decisionmaking process. *Opinion of the Data Protection Working Party on Purpose Limitation*, (Apr. 2, 2013),

given that even where privacy constitutes a fundamental human right, it is not an "*über-value*" that trumps every other social consideration.

This Essay does not provide a comprehensive taxonomy of big data benefits. It would be pretentious to do so, ranking the relative importance of weighty social goals. Rather it posits that such benefits must be accounted for by rigorous analysis taking into account the priorities of a nation, society, or culture. Only then can benefits be assessed *within* the privacy framework.

Consider the following examples of countervailing values (i.e., big data benefits) as they are addressed, with little analytical rigor, by privacy regulators. For example, despite intense pushback from privacy advocates, legislative frameworks all over the world give national security precedence over privacy considerations.[23] On the other hand, although mandated by corporate governance legislation in the United States, whistleblower hotlines are not viewed by privacy regulators as worthy of deference.

What is the doctrinal basis for accepting national security as a benefit that legitimizes privacy costs, while denying the same status to corporate governance laws? Such selective, apparently capricious enforcement is detrimental for privacy. Regulators should pursue a more coherent approach, recognizing the benefits of big data as an integral part of the privacy framework through legitimate interest analysis under the European framework or unfairness doctrine applied by the FTC.

F. *Certainty*

The utility function of big data use depends not only on absolute values, but also on the *probability* of any expected benefits and costs. Not every conceivable benefit, even if highly likely, justifies a privacy loss. Legitimate interest analysis should ensure that lack of certainty of expected benefits is a discounting factor when weighing big data value.

A given level of uncertainty may weigh differently depending on the risk profile of a given culture or society. The United States, for example, established by explorers who pushed the frontier in a lawless atmosphere, continues to highly reward entrepreneurship, innovation, research, and discovery. The quintessential American hero is the lone entrepreneur who against all odds weaves straw into gold. This environment may—and to this day in fact does—endorse practically unfettered data innovation, except in certain regulated areas such as health and financial information, or in cases of demonstrable harm. Failure is considered valuable experience and entrepreneurs may be funded many times over despite unsuccessful outcomes. Conversely, in Europe, the departure point is diametrically opposite, with data processing being prohibited unless a legitimate legal basis is shown.

---

available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

23. *See, e.g.*, Data Protection Act, 1998, c. 29, § 28 (U.K.).

To critics on either side of the Atlantic, both the U.S. and E.U. approaches have their shortcomings. Taken to their extremes, the E.U. approach, with its risk aversion and regulatory bureaucracy, could stifle innovation and growth of a vibrant technology sector, while the U.S. approach, with its *laissez faire* ideology, risks a rude awakening to a reality of eerie surveillance and technological determinism.

CONCLUSION

This symposium issue sets the stage for a discussion of big data that recognizes the weighty considerations on both sides of the value scale. The authors deploy different lenses to expose diverse aspects of the big data privacy conundrum. Some authors focus on the macro, debating broad societal effects: Cynthia Dwork and Deirdre Mulligan discuss the impact of big data on classification, discrimination, and social stratification.[24] Neil Richards and Jonathan King uncover three paradoxes underlying the power structure of the big data ecosystem.[25] Joseph Jerome warns that big data may be socially regressive, potentially exacerbating class disparities.[26] Jonas Lerman examines the overlooked costs of being excluded from big data analysis, suffered by "[b]illions of people worldwide [who] remain on big data's periphery."[27] Ian Kerr and Jessica Earle focus on big data's "preemptive predictions," which could reverse the presumption of innocence, upending the power relationships between government and individuals.[28] Other authors concentrate on the micro, focusing on interpersonal relationships in a data-rich environment: Karen Levy argues that big data has transcended the scope of organizational behavior, entering the delicate domain of individual relationships.[29] Woodrow Hartzog and Evan Selinger predict that absent a robust concept of obscurity, the "data-fication" of personal relationships would strain the social fabric.[30] Other authors seek to harness technology to tame big data effects. Jonathan Mayer and Arvind Narayanan advocate privacy enhancing technologies.[31] Ryan Calo supports organizational measures, such as "consumer subject review boards."[32] Yianni Lagos

---

24. Cynthia Dwork & Deirdre K. Mulligan, *It's Not Privacy, and It's Not Fair*, 66 STAN. L. REV. ONLINE 35 (2013).

25. Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41 (2013).

26. Joseph W. Jerome, *Buying and Selling Privacy: Big Data's Different Burdens and Benefits*, 66 STAN. L. REV. ONLINE 47 (2013).

27. Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55 (2013).

28. Ian Kerr & Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, 66 STAN. L. REV. ONLINE 65 (2013).

29. Karen E.C. Levy, *Relational Big Data*, 66 STAN. L. REV. ONLINE 73 (2013).

30. Woodrow Hartzog & Evan Selinger, *Big Data in Small Hands*, 66 STAN. L. REV. ONLINE 81 (2013).

31. Jonathan Mayer & Arvind Narayanan, *Privacy Substitutes: A Thought Experiment*, 66 STAN. L. REV. ONLINE 89 (2013).

32. Ryan Calo, *Consumer Subject Review Boards*, 66 STAN. L. REV. ONLINE 97 (2013).

and Jules Polonetsky stress the importance of a combination of technological and organizational mechanisms to achieve robust de-identification.[33] We hope that the following essays shift the discussion to a more nuanced, balanced analysis of the fateful value choices at hand.

33. Yianni Lagos & Jules Polonetsky, *Public vs. Nonpublic Data: The Benefits of Administrative Controls*, 66 STAN. L. REV. ONLINE 103 (2013).