Privacy and Big Data: Making Ends Meet

Jules Polonetsky and Omer Tene*

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 story crystalized privacy advocates' concerns of "sleepwalking into a surveillance society" even as decision-makers remain loath to curb government powers for fear of destructive terrorism or cybersecurity attacks.

Over the past few years, the volume of data collected and processed by business and government organizations has increased exponentially. This trend, called "Big Data", is driven by reduced costs of storing information and moving it around in conjunction with increased capacity to instantly analyze massive troves of unstructured data by using modern analytics methods and large-scale statistical simulations. Big Data creates tremendous value 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, over criminalization, and other restricted freedoms.

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 and 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]

---

* 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, THE 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, THE GUARDIAN, June 7, 2013, http://www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data.

[2] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239 (2013); Ira Rubinstein, *Big Data: The End of Privacy or a New Beginning?* 3 INT'L DATA PRIV. L. 74, 78 (2012).

[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; and this is where privacy costs need to be balanced against Big Data benefits. *See* Ann Cavoukian, *Privacy by Design: The Seven Foundational Principles*, January 2011, http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf (noting: "Privacy by Design seeks

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. Privacy harms are notoriously difficult to quantify. As Dan Solove recently notes, "it is very difficult at the time of data collection for a person to make a sensible judgment about the future privacy implications because the implications are often unknown."[4] Consider the use of personal data by merchants to effect price discrimination.[5] The privacy impact of price discrimination is not easy to discern given that one consumer's gain (being charged less for a product or service) is another's loss (being charged more); and that societal gain (more efficient resource allocation) may be offset by harm to broader social goals (inequality and marginalization of weakened groups). Despite this uncertainty, detailed frameworks have developed to help decision-makers understand and quantify privacy risks, with privacy impact assessments (PIA) now increasingly commonplace for government and business undertakings.[6]

However, we argue that accounting for *costs* is only a 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.[7] But the societal value of these benefits may depend on their

---

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] Daniel Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013).

[5] Jennifer Valentino-Devries, Jeremy Singer-Vine & Ashkan Soltani, Websites Vary Prices, Deals Based on Users' Information, WSJ, December 24, 2012, http://online.wsj.com/article_email/SB10001424127887323777204578189391813881534-lMyQjAxMTAyMDIwNDEyNDQyWj.html#12. *See* discussion in Omer Tene & Jules Polonetsky, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, ___ J. TELECOM. & HIGH TECH. L. ___ (forthcoming 2013).

[6] *See, e.g.*, DAVID WRIGHT & PAUL DE HERT (Eds.), PRIVACY IMPACT ASSESSMENT (Springer 2012); Department of Homeland Security, PRIVACY IMPACT ASSESSMENTS: THE PRIVACY OFFICE OFFICIAL GUIDANCE (June 2010), http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_guidance_june2010.pdf.

[7] *See, e.g.*, RICK SMOLAN & JENNIFER ERWITT, THE HUMAN FACE OF BIG DATA (Against All Odds Productions 2012); VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (Eamon Dolan/Houghton Mifflin 2013); World Economic Forum, UNLOCKING THE VALUE OF PERSONAL DATA: FROM COLLECTION TO USAGE, February 2013, http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf; Center for Information Policy Leadership, BIG DATA AND ANALYTICS: SEEKING FOUNDATIONS FOR EFFECTIVE PRIVACY GUIDANCE, February 2013,

nature; on whether they are certain or speculative; and on whether they flow to individuals, communities, businesses, or society at large. Furthermore, existing descriptions of the value of Big Data in multiple settings fail to take the extra step of helping practitioners on the ground *link those benefits to privacy costs* in order to weigh one against the other. And to overlay benefits against costs, decision-makers must have at their disposal an analytical framework to assess benefits.

The integration of benefit considerations into privacy analysis is not without basis in current law. In fact, it fits neatly within the existing privacy doctrine under both the FTC's authority to prohibit "unfair trade practices" in the U.S.[8] as well as the "legitimate interests of the controller" clause in the European Union.[9] Over the past few years, the FTC has carefully refocused its jurisdiction under the "unfairness" strain of its Section 5 "unfair or deceptive acts or practices" powers. 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." [10] 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.[11]

In the European Union, organizations are authorized to process personal data without individual 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".[12] Legitimate interest analysis is inexorably linked to an assessment of benefits. In its recent opinion on "purpose limitation", the Article 29 Working Party provided new impetus for re-purposing data where the new purpose is not "incompatible" with that for which the data were collected.[13]

This article proposes parameters for a newly conceptualized cost-benefit equation, which 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 what is the level of *certainty* that those benefits can be realized. In doing so, it offers ways to introduce into legitimate interest analysis Big Data benefits that accrue not only to businesses but also to individuals and to society at large.

---

http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf; Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63 (2012).

[8] 15 U.S.C. § 45(a)(1).

[9] Article 7(f) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281) 31 (Nov. 23, 1995) http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF (the "Data Protection Directive").

[10] 15 U.S.C. § 45(n) (emphasis added).

[11] Woodrow Hartzog & Daniel Solove, The FTC and the New Common Law of Privacy (unpublished manuscript; on file with authors).

[12] Article 14(a) of the Data Protection Directive.

[13] Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation, WP 203, April 2, 2013, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

Beneficiaries

Who benefits from Big Data? In examining the value of Big Data, we start by evaluating who is affected by the relevant benefit. 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.

Individuals. In certain cases, Big Data analysis provides direct benefit to those individuals whose information is being used, providing 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 decision-making processes.[14] In some cases, a particular practice may be difficult to explain within the brief opportunity that an individual pays attention; in other, 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 pro-actively monitor its customers' computers to detect malware;[15] 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.[16] 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 of proponents of behavioral advertising have had in persuading regulators that personalized ads provide direct benefits to individuals. Behavioral ads are delivered 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.

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

---

[14] 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 (2012).
[15] Roy Furchgott, Comcast to Protect Customer's Computers from Malware, NY TIMES, September 30, 2010, http://gadgetwise.blogs.nytimes.com/2010/09/30/comcast-to-monitor-customer-computers-for-malware.
[16] Daniel Lippman & Julian Barnes, Malware Threat to Internet Corralled, WSJ, July 9, 2012, http://online.wsj.com/article/SB10001424052702303292204577515262710139518.html.

similar product or residents in a certain 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 not only individually but also to other users and prospective drug users.[17]

Organizations. Big Data analysis often benefits the businesses or other organizations that collect and harness the data. Data-driven profits may be viewed as enhancing allocative efficiency by facilitating the "free" economy.[18] 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 [19] and fermenting social change. [20] With more data, businesses can optimize distribution methods,[21] efficiently allocate credit and robustly combat fraud, benefitting consumers as a whole. But in the absence of individual value or broader societal gain, others may consider enhanced business profits to be a 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.

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 thereby 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.

Benefits

Big Data creates enormous value for the global economy, driving innovation, productivity, efficiency and growth. The uses of Big Data can be transformative and are sometimes difficult to anticipate at the time of initial collection. Data has become the driving force behind almost every interaction between individuals, businesses and governments. Each category of benefits carries different value to different audiences or cultures.

Society must come up with criteria to evaluate the relative weight it gives different benefits – or social values – in order to allow privacy regulators to assess whether in a

---

[17] Nicholas Tatonetti et al., *A Novel Signal Detection Algorithm for Identifying Hidden Drug-Drug Interactions in Adverse Event Reports*, 12 J. AM. MED. INFORMATICS ASS'N 79, 79–80 (2011).
[18] CHRIS ANDERSON, FREE: THE FUTURE OF A RADICAL PRICE (2009).
[19] Tim Worstall, UN Report: More People Have Mobile Phones Than Toilets, FORBES, March 23, 2013, http://www.forbes.com/sites/timworstall/2013/03/23/more-people-have-mobile-phones-than-toilets.
[20] WAEL GHONIM, REVOLUTION 2.0: THE POWER OF THE PEOPLE IS GREATER THAN THE PEOPLE IN POWER: A MEMOIR (Houghton Mifflin Harcourt 2012).
[21] *A Different Game: Information is Transforming Traditional Businesses*, THE ECONOMIST, Feb. 25, 2010, http://www.economist.com/node/15557465.

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

This article does not provide a comprehensive taxonomy of Big Data benefits. 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 U.S., whistleblowing hotlines are not viewed by privacy regulators as worthy of deference. Another example concerns Google driving through cities all over the world to create a comprehensive map of Wi-Fi networks for its geo-location services. The decisions by regulators in this case indicate some appreciation for the value created by Google, even if this rationale was not clearly expressed.

What is the doctrinal basis for accepting national security and geo-location mapping as benefits that legitimize privacy costs while denying the same status to U.S. corporate governance laws? This track record of selective enforcement is detrimental for privacy. Regulators should pursue a more nuanced 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.

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 unlikely, 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 risk or uncertainty may weigh differently depending on the risk profile of differing societies. The U.S., 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

[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 decision-making process.

[23] *See, e.g.*, Section 28 of the UK Data Protection Act, 1998, Data Protection Act 1998, c. 29.

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 despite unsuccessful outcomes. Conversely, in Europe, the departure point is diametrically opposite, with data processing is prohibited unless a "legitimate legal basis" is shown.

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

Conclusion

This symposium issue sets the stage for a discussion of Big Data that recognizes the weighty values on both sides of the scale. We hope that the following essays shift the discussion to a more balanced, nuanced analysis of the fateful value choices at hand.