Big data analytics and the right to privacy

Jenifer Sunrise Winter

My work addresses communication rights in the context of the widespread growth and expansion of the Internet. In particular, I focus on privacy rights in the context of Internet-based data collection and analysis (big data analytics).

Privacy is explicitly stated under Article 12 of the Universal Declaration of Human Rights and is seen as an enabler of other communication rights, such as freedom of expression and freedom of association. However, with the widespread diffusion of the Internet, technical barriers to surveillance have diminished, and legal and regulatory protections for personal privacy have been unable to address the complexity of related technical and economic changes. Because privacy is an abstract, seemingly intangible concept, deeply interwoven with other rights and dependent on context, it is often difficult to define and assess.

My task as a scholar is to document instances where citizens feel that their personal information has been inappropriately collected, used, or shared, and then (along with the work of other scholars) map out the benefits and harms of the technological systems that are enabling surveillance. I seek to provide evidence about how big data analytics can expose private information, thereby disadvantaging certain individuals or groups, creating unjust power differentials, and hindering public participation in democratic discourse – weighing this against the many possible benefits.

While we tend to think of the Internet as something virtual and intangible, information accessed via communication devices such as a smartphone or mobile personal computer, it is rapidly becoming part of the natural world itself. An increasing number of ordinary objects are being designed, or redesigned, to include digital sensors, computing power, and communication capabilities.

By 2020, an estimated 50-75 billion "Things" (intelligent, everyday objects) may be connected to the Internet (Danova, 2013). As more of these become present in our daily environment and communicate over the Internet, the "real" and virtual worlds are merging. This so-called "Internet of Things" is poised to offer many benefits, but it also poses many threats to personal privacy and self-expression.

More data is expected to flow over the

Internet in 2016 than in the sum of all previous years since 1969 (Cisco, 2014). Data are increasingly generated and analyzed by computers, such as those embedded in modern automobiles, smart meters measuring electrical use in the home, and appliances in our homes. Many of these data are *personal* – for example, televisions or gaming consoles might record all of the audio in a room, capturing our personal conversations, or they may use facial-recognition features that observe when one is watching a particular television program or playing a certain game.

For example, Vizio smart televisions record the date, time, and channel of programs you watch and sell this information, along with your IP address, to data brokers such as Neustar, who combine it with hundreds of personal attributes

(e.g., your age, marital status, estimated wealth, credit score). Increasingly, data from smart "Things" is being combined with data from public records (e.g., address, property taxes, criminal records, divorce proceedings, and demographics, as well as data shared via social network sites).

Stripping away anonymity

Big data analytics harnessed by the emerging Internet of Things undoubtedly has the potential to offer a great deal of societal benefit; however, there is increasing evidence of unjust discrimination for some individuals and groups. In some jurisdictions, laws require that sensitive data are stripped of personally-identifiable information, anonymizing them to avoid harm. However, the volume of data now collected, and the sophistication of the tools used to mine it, makes it possible to re-identify anonymous data.

Medical researchers matched personal DNA sequences shared on Internet genealogy forums with other publicly available data – uniquely identifying many subjects (Gymrek, McGuire, Golan, Halperin & Erlich, 2013). Location-based data, such as that collected by your phone throughout the day, can also uniquely identify you. Researchers examining vehicle location data as part of the Telecom Italia Big Data Challenge, found that only a handful of data points (locations) was needed to uniquely identify drivers in the city of Milan, and that there is no way, at present, to anonymize these data (Manfredi, Mir, Lu, & Sanchez, 2014).

Anonymity is essential to democratic political practice, as it provides freedom to seek information and express ideas – however unpopular – including those critical of government (Solove, 2011). Anonymity enables citizens to go about their everyday lives without inhibition, and therefore its absence can stifle dissent and free expression. As citizens have become increasingly aware of state and corporate surveillance, they may self-censor, avoiding seeking information, sharing ideas, or associating with others due to fear of surveillance and reprisal.

A 2013 study by a national group representing journalists, cartoonists, and other writers, surveyed its members and found that they reported increasingly engaging in self-censorship due to growing awareness of government mass surveillance programs that monitor the activities of everyday citizens (PEN American Center, 2013).

Impact on individuals

Analysis of large data sets may reveal patterns that allow governments or marketers to infer certain things about a person, or even forecast his or her behavior. As more and more data are being collected related to everyday tasks that seem innocuous – location over time, items purchased or browsed for, entertainment media consumed, "friends" added or looked at on social networks, or proximity with certain people or places, they can be used to make judgements that affect an individual and his or her life chances.

For example, in two recent studies I conducted, participants expressed concern that analysis of these data may lead to discrimination when individuals seek housing, immigration eligibility, or employment (Winter, 2014; Winter, 2015). Access to private data creates an asymmetry between citizens, who are surveilled and categorized, and states or corporations, who can mine it for insight.

For example, insurance companies may no longer spread risk across a large group when they are able to cut off those deemed less profitable based on big data analytics. Upturn (2014) notes that this is already occurring, as the collection of "non-traditional" third-party data sources enables prediction that is just as accurate as a medical examination. This may lead to higher financial burdens for those with certain medical conditions, as well as for healthy individuals categorized as high-risk due to late-night driving or living in a low-income area, groups already populated by vulnerable populations.

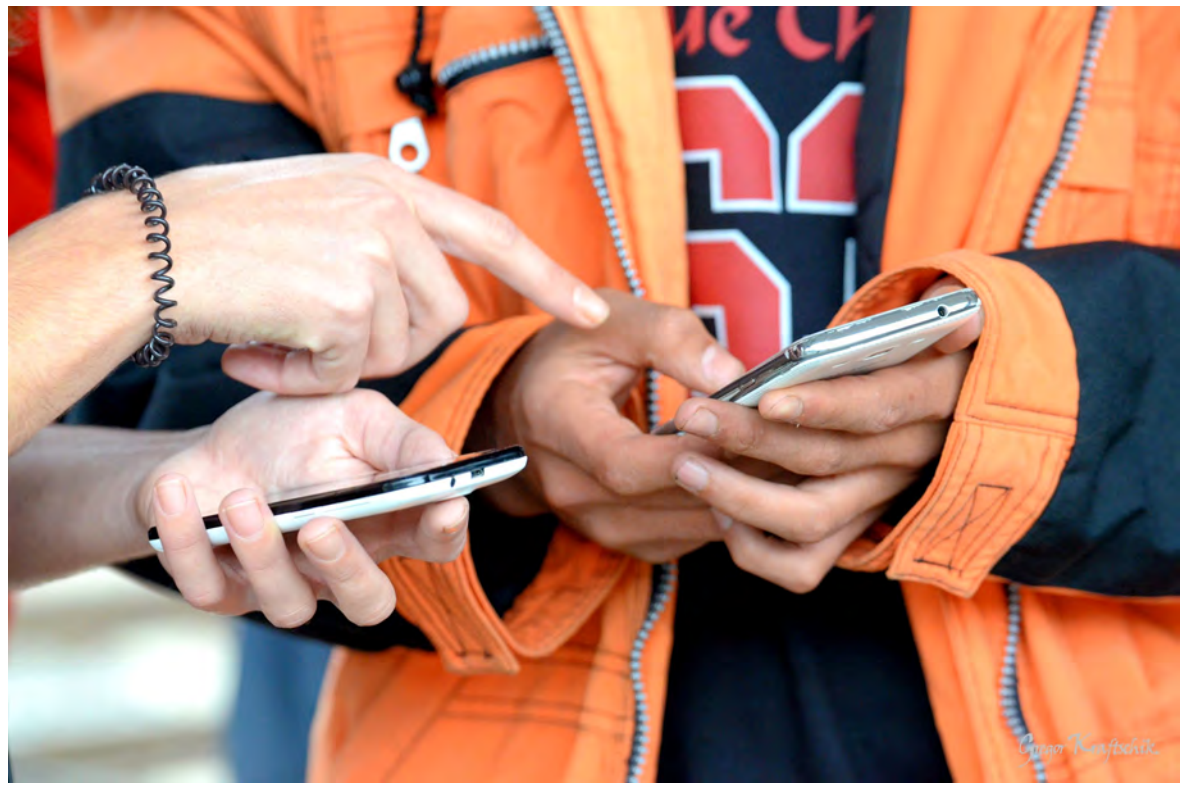
Similarly, when you bring your excess change to a self-service coin-counting kiosk, banks may be purchasing that data and using it to estimate the probability that you will default on your mortgage (i.e., they may assume that if you are constantly turning in coins, you may be struggling financially). In many cases, data collection

lacks transparency, so citizens are unaware of what is being collected, how it is being analyzed, or with whom it is being shared.

Despite legal protections, there may be instances where the monetary value of a data set outweighs an organization's concern with ethical or legal restrictions. Energy use data gathered by smart meters in the home, for example, might reveal specific lifestyle information that could be used by insurers or commercial service providers (National Institute of Standards and Technology, 2010). Genetic information, increasingly available in anonymized form on the Internet, is another area of concern. Despite laws such as the *Genetic Information Nondiscrimination Act* (United States), there are powerful financial incentives for mining. Where such laws restrict use, big data analytics may simply rely on non-protected "proxy" fields, allowing other, non-protected information that correlates with the variable of interest to be used instead. Barocas and Selbst (2016) explain how process-oriented civil rights laws cannot adequately address the disparate impact of big data analytics, as discrimination based on proxies will often "discover" patterns that continue to reinforce existing social inequalities.

It should be noted that big data analytics is not limited to countries and regions traditionally rich in information and communication technologies – it is also poised to transform developing nations. Even where the Internet is not yet robust, the growth of mobile networks has enabled big data analytics. Credit assessment has been revolutionized by the predictive analysis of "non-traditional" data gathered via mobile phone use. On the positive side, this is bringing affordable credit to "hundreds of millions of aspiring middle-class consumers in emerging markets" in nations such as Kenya and Columbia (Costa, Deb, & Kubzansky, 2015, p. 4).

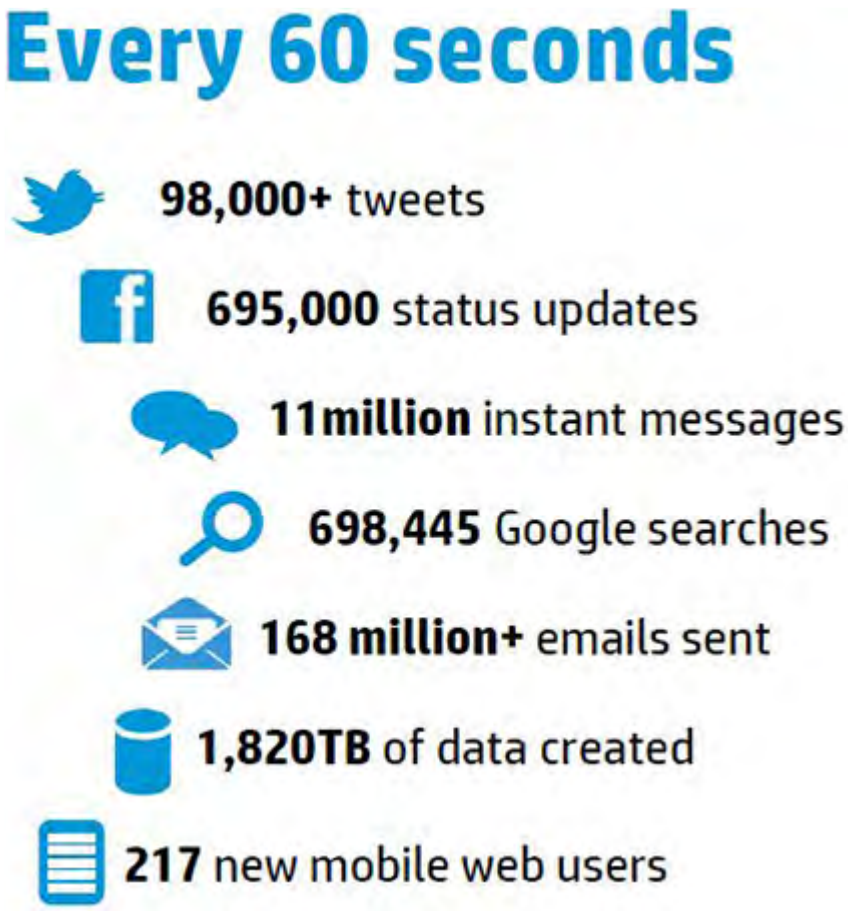
A similar positive benefit of big data analytics has been reported in regard to the Middle



East refugee crisis. As reported by Marr (2015), Nagina Kaur Dhanoa, CIO for the United Nations High Commission for Refugees, observed that "The first thing people running the Za'atri [refugee] camp in Jordan ask for is not tents and blankets, but where they can charge their mobile phone" (para. 4). Mobiles help refugees connect to health services, and locate food, supplies, and housing. They also allow governments, NGOs, and corporations to create unique data profiles for each refugee, creating the possibility for personal data abuse.

As big data analytics continues to penetrate the core of business and governmental decision-making around the world, my research seeks to assess the distribution of both benefits and harms resulting from data-gathering activities, as well as to explore whether the distributions that result are in conflict with general moral and political principles related to community values.

For example, in many democracies, privacy is understood as a concept that supports democratic values (e.g., equality and justice). By cataloging these distributions, scholars can influence policymakers. The results of two of my studies (along with other scholars' work) addressing privacy concerns related to the Internet of Things and smart meters have been recently included in public comments before the Federal Trade Commission, a United States government agency tasked with consumer protection.



Further, because big data analytics and the emerging Internet of Things invariably put some subset(s) of citizens at risk, I have sought to make the process itself more transparent to citizens by taking part in media interviews and creating accessible summaries of research articles for the public.

While the effects of these efforts are not immediately observable, an accumulation of such research by scholars has led to governments around the world strengthening personal data protection regulations and has brought critical issues related to privacy and freedom of expression to the attention of the public, fostering informed debate. ■

Image credits: Page 10 "Smile, you are on camera," by Intel Free Press under CC BY-SA 2.0. Page 12 "A refugee helper and a refugee overcome language barriers by using their mobile phones," by G. Kraftschik under CC-BY. Page 13: Borja Burguillos.

References

Angwin, J. (2015, November 9). Own a Vizio smart TV? It's watching you. Pro Publica. Retrieved from https://www.propublica.org/article/own-a-vizio-smart-tv-its-watching-you

Barocas, S., & Selbst, A.D. (2016). Big data's disparate impact. *California Law Review, 104.*

Cisco. (2014, June 10). *The Zettabyte era: Trends and analysis.* Cisco White Paper. San Jose, Ca.: Cisco Systems. Retrieved from http://www.cisco.com/c/en/us/ solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.pdf

Costa, A., Deb, A., & Kubzansky, M. (2015). Big data, small credit: The digital revolution and its impact on emerging market consumers. Redwood City, Ca.: Omidyar network.

Danova, T. (2013, October 2). "Morgan Stanley: 75 Billion Devices Will Be Connected To The Internet Of Things By 2020." Retrieved from http://www.businessinsider.com/75-billion-devices-will-be-connected-to-the-internet-by-2020-2013-10

Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. Science, 339(6117), 321-324.

Manfredi, N., Mir, D., Lu, S., & Sanchez, D. (2014). Differentially private models of tollgate usage: The Milan tollgate data set. *IEEE International Conference on Big Data, 2014*, 46-48.

Marr, B. (2015, October 15). "Big data, technology, and the Middle East refugee crisis." Forbes. Retrieved from http://www.forbes.com/sites/bernardmarr/2015/10/15/big-data-technology-and-the-middle-east-refugee-crisis/

National Institute of Standards and Technology. (2010). Introduction to NISTIR 7628: Guidelines for smart grid cyber security. Gaithersburg, MD: NIST.

PEN American Center. (2013). *Chilling effects: NSA surveillance drives U.S. writers to self-censor.* New York. PEN American Center.

Schwartz, P.M., & Solove, D. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review, 86*, 1814-1894.

Winter, J.S. (2015). Citizen perspectives on the customization/privacy paradox related to smart meter implementation. *International Journal of Technoethics, 6*(1), 45-59.

Winter, J.S. (2014). Surveillance in ubiquitous network societies: Normative conflicts related to the consumer in-store supermarket experience in the context of the Internet of Things. *Ethics and Information Technology, 16*(1), 27-41.

Jenifer Sunrise Winter is Associate Professor in the School of Communications at the University of Hawai'i at M□noa. Her research focuses on communication rights, in particular privacy and the Internet of Things. Related research addresses broadband access rights, and the Internet as a support of democratic institutions and publics. She serves as Secretary of the Right to Communicate Group.