Introduction to the Special Issue: Digital Inequalities and Discrimination in the Big Data Era
Author(s): Jenifer Sunrise Winter
Source: *Journal of Information Policy*, Vol. 8 (2018), pp. 1-4
Published by: Penn State University Press
Stable URL: http://www.jstor.org/stable/10.5325/jinfopoli.8.2018.0001
Accessed: 10-05-2018 17:37 UTC

INTRODUCTION TO THE SPECIAL ISSUE

Digital Inequalities and Discrimination in the Big Data Era

Jenifer Sunrise Winter

A growing number of ordinary objects are being redesigned to include digital sensors, computing power, and communication capabilities—and new objects and processes are becoming part of the Internet. This emerging Internet of Things (IoT) ecosystem—networks of physical objects embedded with the ability to sense, and sometimes act upon, their environment, as well as related communication, applications, and data analysis—enables data to be collected from billions of everyday objects. The emerging datasphere made possible by these developments offers immense potential to serve the public good by fostering government transparency, energy conservation, participatory governance, and substantial advances in medical research and care. On the other hand, a growing body of research addresses emerging privacy and civil liberties concerns related to big data, including unjust discrimination and unequal access to data and the tools needed to make use of it.

For example, big data analytics may reveal behavioral or psychographic patterns that were previously not detectable. Data about a variety of daily tasks that seem trivial are increasingly aggregated and used to reveal associations or behaviors, and these analyses can be used to make judgements that harm individuals or groups. With the advent of artificial intelligence and machine learning, many decisions around us are shaped by these calculations rather than traditional human judgement. For example, sensitive personal information or behaviors (e.g., political or health related) may be used to discriminate when individuals seek housing, immigration eligibility, medical care, education, bank loans or other financial services, insurance, or employment.

Jenifer Sunrise Winter: School of Communications, Pacific Information and Communication Technology for Development Collaborative, University of Hawai'i at Mānoa

JOURNAL OF INFORMATION POLICY, Volume 8, 2018

Governing the collection and use of personally identifiable data is proving a great challenge in the Big Data Era. We are confronted with a lack of transparency about what data are collected, how they are being used, who is deriving value from their use (e.g., individuals, society at large, and corporate interests), and what algorithms are being used to make decisions—what Frank Pasquale has called the "Black Box Society."[1] Even where efforts are made to avoid discrimination based on protected attributes, big data analytics may enable "proxy" variables that highly correlate to the protected variable. For example, Solon Barocas and Andrew Selbst have demonstrated how process-oriented civil rights laws (e.g., those seeking to protect against racial or gender discrimination) cannot adequately address big data's disparate impact, as discrimination based on these proxies is prone to "discover" patterns that reinforce existing social inequalities.[2] Whether existing regulatory efforts and institutions can evolve to address these critical challenges, or whether novel forms of intervention will be required, is emerging as a key policy concern in the United States and internationally.

In May 2017, the Pacific Information and Communication Technology for Development Collaborative (PICTDC) at the University of Hawai'i at Mānoa and the Institute for Information Policy (IIP) at Pennsylvania State University cosponsored a preconference at the International Communication Association Annual Conference in San Diego to address these themes. Following peer review, three of those articles are published here.

In the first article, "A Proposal to Adopt Data Discrimination Rather Than Privacy as the Justification for Rolling Back Data Surveillance," Benjamin Cramer argues that legal strategies based on privacy have failed in the United States to convince judges or politicians to limit surveillance enabled by big data analytics. Privacy, he notes, is still a high-level ideal, but it is a vague concept and an implied right, rather than one explicitly noted in the Constitution. Existing US privacy statutes are unable to address the pervasive surveillance enabled by big data. Instead, he argues that we should explore the use of unjust data discrimination in legal arguments seeking to limit surveillance, looking to statutory and judicial precedents related to other types of discrimination.[3]

1. Pasquale.
2. Barocas and Selbst.
3. Cramer.

In the second article, "Health Wearables: Ensuring Fairness, Preventing Discrimination, and Promoting Equity in an Emerging Internet-of-Things Environment," Kathryn Montgomery, Jeff Chester, and Katharina Kopp introduce key trends that are influencing the growth of a personally generated health data marketplace linked to wearable health devices. While acknowledging the potential for substantial public health improvements, they describe possible risks of this big data ecosystem to both individuals and broader society, such as the aggregation of personal health information with other data sources, leading to potential discriminatory profiling, manipulative marketing, and security breaches. For example, they note how predictive analytics have led to means for "scoring" and categorizing individuals based on these data, and even estimating the probability that one will develop certain health conditions based on comparison with others' personal data. To address these substantial challenges, the authors assess present regulatory frameworks and self-regulatory regimes and highlight key principles and issues that must be addressed in order to establish a public interest framework for the health wearables data marketplace.[4]

In the third article, "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications," Betsy Anne Williams, Catherine Brooks, and Yotam Shmargad address how data-driven models used to inform important decisions in a variety of institutional decision-making processes can discriminate even where no explicit social categories have been collected (e.g., gender or race) through proxy variables that have strong correlation to the variables of interest. They provide illuminating examples of how automated decision-making can be biased, and how these biases are then amplified and reinforced via data analytics. They demonstrate how a lack of social categories may even exacerbate existing biases by making them harder to detect and address. They argue—paradoxically—that institutions should collect and carefully use social category data in order to identify these biases. They offer four illustrative case studies, each with relevant remedy: using data for an external audit; detecting and eliminating biased standardized test questions; understanding how legal, ethical, and social concerns shape policies and outcomes; and hiring individuals from underrepresented groups in Science, Technology, Engineering, and Math (STEM) jobs.[5]

4. Montgomery, Chester, and Kopp.

5. Williams, Brooks, and Shmargad.

This *Journal of Information Policy* Special Issue on Digital Inequalities and Discrimination in the Big Data Era is intended to stimulate further academic research and policy discourse about how to govern personally identifiable information in the Big Data Era. How can scholars continue to illuminate the unjust discriminatory impacts of big data analytics, particularly given the lack of transparency in these interactions? What role do system designers and corporations play in addressing these concerns? How can policy makers address the array of related challenges? Are existing institutions able to evolve to better address these monumental shifts? A new vision is required to address these and other questions. The articles chosen for this special issue demonstrate how rigorous academic research can impact and address real-world problems and contribute to policy solutions that serve the public's best interest.

BIBLIOGRAPHY

Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review* 104 (2016): 671–732.

Cramer, Benjamin W. "A Proposal to Adopt Data Discrimination Rather Than Privacy as the Justification for Rolling Back Data Surveillance." *The Journal of Information Policy* 8 (February 2018): 5–33.

Montgomery, Kathryn, Jeff Chester, and Katharina Kopp. "Health Wearables: Ensuring Fairness, Preventing Discrimination, and Promoting Equity in an Emerging Internet-of-Things Environment." *The Journal of Information Policy* 8 (February 2018): 34–77.

Pasquale, Frank. *The Black Box Society*. Cambridge, MA: Harvard University Press, 2015.

Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications." *The Journal of Information Policy* 8 (February 2018): 78–115.