It's Not Privacy, and it's Not Fair

Cynthia Dwork & Deirdre K. Mulligan

Classification is the foundation of targeting and tailoring of information and experiences to individuals. Big data promises—or threatens—to bring classification to an increasing range of human activity. While many companies foster an illusion that classification is an area of absolute algorithmic rule—that decisions are neutral, organic, and automatically rendered without human intervention—reality is a far messier mix of technical and human curation. Both the data set and the algorithms reflect choices about data, connections, inferences, interpretation, thresholds for inclusion, etc., that advance a specific purpose. Like maps that represent the physical environment in varied ways to serve different needs—mountaineering, sightseeing, shopping—classification systems are neither neutral nor objective, but are biased toward their purpose. They reflect the explicit and implicit values of their designers. Yet, few designers "see them as artifacts embodying moral and aesthetic choices" or recognize the powerful role they play in crafting "people's identities, aspirations and dignity."[1] But increasingly, the subjects of classification, as well as regulators, do.

Today, the creation and consequences of some classification systems, such as systems for online behavioral advertising (OBA), are under scrutiny by consumer and data protection regulators, and advocacy organizations. Every step in the big data pipeline of OBA is raising concerns: from the privacy implications of amassing, connecting, and using personal information; to the implicit and explicit biases embedded in both data sets and algorithms; to the individual and societal consequences of the resulting classifications and segmentation. Although the concerns are wide ranging and complex, the discussion and proposed solutions inevitably loop back to privacy and transparency: specifically, establishing

[1] Bowker, Geoffrey C. & Star, Susan Leigh, SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES. p. 4. The MIT Press 2000.

individual control over personal information, and requiring entities to provide some transparency into personal profiles and algorithms.[2]

The computer science community, while acknowledging concerns about discrimination, tends to position privacy as the dominant concern[3]. Privacy-preserving advertising schemes support the view that tracking, auctioning, and optimizing done by the many parties in the advertising ecosystem is acceptable, as long as these parties don't "know" the identity of the target[4].

Policy proposals are similarly narrow. They include regulations requiring consent prior to tracking individuals or prior to the collection of "sensitive information"; and context-specific codes respecting privacy expectations.[5] Bridging the technical and policy arenas, the World Wide Web Consortium's draft "do-not-track" specification will allow users to signal a desire to avoid OBA. Greater transparency is part of these approaches.

Regrettably, privacy controls and increased transparency fail to address concerns with the classifications and segmentation produced by big data analysis.

At best, solutions that vest individuals with control over personal data indirectly impact the fairness of classifications and outcomes—discrimination in the

---

[2] See L. Introna and H. Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society*, 16(3):1-17, 2000; Pasquale, Frank A., "Restoring Transparency to Automated Authority," Seton Hall Research Paper No. 2010-28; Danielle Keats Citron, "Technological Due Process," 85 Wash. U.L.Rev. 1249, 1308-9 (2008); Daniel J. Steinbock, "Data Matching, Data Mining, and Due Process," 40 Ga. L.Rev. 1, 17 (2005). For an example see, Department of Homeland Security provides access to passenger name record system data but not the decisional rules of the system. DEPARTMENT OF HOMELAND SECURITY Office of the Secretary 6 CFR Part 5 [Docket No. DHS-2009-0055] Privacy Act of 1974: Implementation of Exemptions; Department of Homeland Security/U.S. Customs and Border Protection--006 Automated Targeting System of Records AGENCY: Privacy Office, DHS. ACTION: Final rule.

[3] "Some are concerned that OBA is manipulative and discriminatory, but the dominant concern is its implications for privacy." V. Toubiana, A. Narayana, D. Boneh, H. Nissenbaum, and S. Barocas, "Adnostic: Privacy Preserving Targeted Advertising," NDSS 2010.

[4] "The privacy goals…are…Unlinkability: the broker cannot associate…information with a single (anonymous) client." A. Reznichenko, S. Guha, P. Francis, "Auctions in Do-Not-Track Compliant Internet Advertising.

[5] Article 5(3) of the amended EU e-Privacy Directive (Directive 2002/58/EC as amended by Directive 2009/136/EC); Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers, Federal Trade Commission, March 26, 2012 at 45; and Department of Commerce, National Telecommunications and Information Administration, Multistakeholder Process to Develop Consumer Data Privacy Codes of Conduct, *Federal Register*, Vol. 77, No. 43, March 5, 2012.

narrow legal sense, "cumulative disadvantage"[6] fed by the narrowing of possibilities, or "filter bubbles."[7] Whether the information used for classification is obtained with or without permission is unrelated to the production of disadvantage or discrimination.

At worst, privacy solutions can hinder efforts to identify classifications that unintentionally produce objectionable outcomes—for example differential treatment that tracks race or gender—by limiting the availability of data about such attributes. Protecting against discriminatory impact—as opposed to intent—is advanced by data about legally protected statuses, as detection often turns on statistics.[8] While automated decision-making systems "may reduce the impact of biased individuals, they may also normalize the far more massive impacts of system-level biases and blind spots."[9] Rooting out biases and blind spots in big data depends on our ability to constrain, understand, and test the systems that use it to shape information, experiences, and opportunities. This requires data.

Exposing the data sets and algorithms of big data analysis to scrutiny—transparency solutions—may improve individual comprehension, but given the independent (sometimes, intended) complexity of algorithms it is unreasonable to expect transparency alone to root out bias.

The decreased exposure to differing perspectives, reduced individual autonomy, and loss of serendipity due to classifications that shackle users to the profiles of individuals and groups used to frame their "relevant" experience, are not privacy problems. While narrowcasting and segmentation are fueled by personal data, they don't depend on it. And, individuals often create their own bubbles. Allowing individuals to peel back their bubbles—to view the Web from someone else's perspective, or devoid of personalization—does not force them outside their

---

[6] Oscar H. Gandy, "Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems," Ethics Inf. Technol (2010) 12:29-42.

[7] ELI PARISER, THE FILTER BUBBLE: HOW THE NEW PERSONALIZED WEB IS CHANGING WHAT WE READ AND HOW WE THINK (2011).

[8] Julie Ringelheim, *op. cit.,* at 14 (discussing the use of demographic data to identify disparate impact in "neutral" rules).

[9] Gandy at

bubbles. [10]

Solutions to these problems are among the hardest to conceptualize, in part because perfecting individual choice may impair socially desirable outcomes. Fragmentation, regardless of whether its impact can be viewed as disadvantageous from any individual's or group's perspective, and whether it is chosen or imposed, corrodes the public deliberation and debate considered essential to a functioning democracy.

If privacy and transparency are not the panacea to the risks posed by big data, what is?

First, we must carefully unpack and model the problems attributed to big data.[11] The ease with which policy and technical proposals revert to solutions focused on individual control over personal information reflects a failure to accurately conceptualize other concerns. While proposed solutions are responsive to a subset of privacy concerns—we discuss other concepts of privacy at risk in big data in a separate paper—they offer a mixed bag with respect to discrimination, and are not responsive to concerns about the ills that segmentation portends for the public sphere.

Second, we must approach big data as a socio-technical system. Objections to automated decision-making have been around as long as information systems. European law generally prohibits decisions based on automated processing of personal information absent human review. In other areas automated-decision making is exalted as the antidote to the discriminatory urges and intuitions of people.[12] Viewing the problem as one of machine versus man is a barrier to addressing concerns with bias in what are mixed socio-technical systems. The key

---

[10] See L"Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society*; "Restoring Transparency to Automated Authority," Seton Hall Research Paper No. 2010-28.

[11] Recent symposia have begun this, see, Governing Algorithms: A conference on computation, automation and control, May 16-17, 2013, New York University; and, Technology: Transforming the Regulatory Endeavor, March 3, 2011, Berkeley Center for Law & Technology, University of California at Berkeley.

[12] For example, see, Fed. Fin. Insts. Examination Council, Interagency Fair Lending Examination Procedures 8 (2007), pp 7-9.

lies in thinking about how best to manage the risks to the values at stake.[13] Questions of oversight and accountability should inform the decision of where to locate values. Code presents challenges to oversight, but policies amenable to formal description can be built in and tested for. The same cannot be said of the brain. Our point is simply that big data debates are ultimately about values first, and only second about math and machines.

Third, lawyers and technologists must focus their attention on the risks of segmentation inherent in classification. There is a broad literature on fairness, notably in social choice theory, game theory, economics, and law[14] that can guide such work. Policy solutions found in other areas include: the creation of "standard offers"; the use of test files to identify biased outputs based on ostensibly unbiased inputs; required disclosures of categories, classes, inputs, and algorithms; and public participation in the design and review of systems used by governments.
In computer science and statistics, the literature addressing bias in classification comprises: testing for statistical evidence of bias; training unbiased classifiers using biased historical data; a statistical approach to situation testing in historical data; a method for maximizing utility subject to any context-specific notion of fairness; an approach to fair affirmative action, and work on learning fair representations with the goal of enabling fair classification of future, not yet seen, individuals.

Drawing from existing approaches, a system could place the task of constructing a metric—defining who must be treated similarly—outside the system, creating a path for external stakeholders—policy makers, others—to have greater influence over, and comfort with, the fairness of classifications. Test files could be used to ensure outcomes comport with the similarity metric. While incomplete, this suggests that there are opportunities to address concerns about discrimination and disadvantage. Combined with greater transparency and individual access rights to data profiles, thoughtful policy and technical design could tend to a more complete set of objections.

---

[13] For example see, Roger Brownsword, "Lost in Translation: Legality, Regulatory Margins, and Technological management," 26 Berkeley Tech. L.J. 1321 (2011.

Finally, the concerns related to fragmentation of the public sphere and "filter bubbles" are a conceptual muddle and an open technical design problem. Issues of selective exposure to media, the absence of serendipity, and yearning for the glue of civic engagement are all relevant. While these objections to classification may seem at odds with "relevance" and personalization, they are not a desire for irrelevance or under-specificity. Rather they reflect a desire for the tumult of traditional public forums—sidewalks, public parks, street corners—where a measure of randomness and unpredictability reigns, yields a mix of discoveries and encounters, that contribute to a more aware and informed populace. They resonate with calls for "public" or "civic" journalism that seeks to engage "citizens in deliberation and problem-solving, as members of larger, politically involved publics" rather than catering to consumers narrowly focused on private lives, consumption, and infotainment. Equally importantly, they reflect the hopes and aspirations we ascribe to algorithms, despite our cynicism and reservations, "we want them to be neutral, we want them to be reliable, we want them to be the effective ways in which we come to know what is most important."[15]

The urge to classify is human; however, the lever of big data brings ubiquitous classification, demanding greater attention to the values embedded and reflected, and the roles they play in shaping public and private life.

---

[15] Tarleton Gillespie, " Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us," Culture Digitally, Oct 19, 2011.