66 STAN. L. REV. ONLINE 35
September 3, 2013

IT'S NOT PRIVACY, AND IT'S NOT FAIR

Cynthia Dwork & Deirdre K. Mulligan*

Classification is the foundation of targeting and tailoring information and experiences to individuals. Big data promises—or threatens—to bring classification to an increasing range of human activity. While many companies and government agencies foster an illusion that classification is (or should be) an area of absolute algorithmic rule—that decisions are neutral, organic, and even automatically rendered without human intervention—reality is a far messier mix of technical and human curating. Both the datasets and the algorithms reflect choices, among others, about data, connections, inferences, interpretation, and thresholds for inclusion that advance a specific purpose. Like maps that represent the physical environment in varied ways to serve different needs—mountaineering, sightseeing, or shopping—classification systems are neither neutral nor objective, but are biased toward their purposes. They reflect the explicit and implicit values of their designers. 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, from determination of tax-exempt status to predictive analytics in health insurance, from targeting for surveillance to systems for online behavioral advertising (OBA), are under scrutiny by consumer and data protection regulators,

* Cynthia Dwork is Distinguished Scientist, Microsoft Research. Deirdre K. Mulligan is Assistant Professor of School of Information, Berkeley Law; Co-Director, Berkeley Center for Law and Technology. The authors would like to thank participants at the Privacy Law Scholars Conference 2012. The project was supported in part by the U.S. Department of Homeland Security under Grant Award Number 2006-CS-001-000001, under the auspices of the Institute for Information Infrastructure Protection (I3P) research program, managed by Dartmouth College. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security, the I3P, or Dartmouth College.

1. GEOFFREY C. BOWKER & SUSAN LEIGH STAR, SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES 4 (2000).

advocacy organizations and even Congress. Every step in the big data pipeline is raising concerns: the privacy implications of amassing, connecting, and using personal information, the implicit and explicit biases embedded in both datasets and algorithms, and the individual and societal consequences of the resulting classifications and segmentation. Although the concerns are wide ranging and complex, the discussion and proposed solutions often loop back to privacy and transparency—specifically, establishing 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 are 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.[6] These approaches involve greater transparency.

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—resulting in discrimination in the narrow legal sense, or "cumulative disadvantage" fed by

---

2. *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1308-09 (2008); Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169 (2000); Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. ON TELECOMM. & HIGH TECH. L. 235 (2011); Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1 (2005).

3. Vincent Toubiana et al., *Adnostic: Privacy Preserving Targeted Advertising* 1 (17th Annual Network & Distributed Sys. Sec. Symposium Whitepaper, 2010), *available at* http://www.isoc.org/isoc/conferences/ndss/10/pdf/05.pdf ("Some are concerned that OBA is manipulative and discriminatory, but the dominant concern is its implications for privacy.").

4. Alexey Reznichenko et al., *Auctions in Do-Not-Track Compliant Internet Advertising*, 18 PROC. ACM CONF. ON COMPUTER & COMM. SECURITY 667, 668 (2011) ("The privacy goals . . . are . . . [u]nlinkability: the broker cannot associate . . . information with a single (anonymous) client.").

5. Multistakeholder Process to Develop Consumer Data Privacy Codes of Conduct, 77 Fed. Reg. 13,098 (Mar. 5, 2012); Council Directive 2009/136, art. 2, 2009 O.J. (L 337) 5 (EC) (amending Council Directive 2002/58, art. 5); FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 45-46 (2012), *available at* http://ftc.gov/os/2012/03/12032privacyreport.pdf.

6. World Wide Web Consortium, *Tracking Preference Expression (DNT), W3C Editor's Draft*, WORLD WIDE WEB CONSORTIUM (June 25, 2013), http://www.w3.org/2011/tracking-protection/drafts/tracking-dnt.html.

the narrowing of possibilities.[7] Whether the information used for classification is obtained with or without permission is unrelated to the production of disadvantage or discrimination. Control-based solutions are a similarly poor response to concerns about the social fragmentation of "filter bubbles"[8] that create feedback loops reaffirming and narrowing individuals' worldviews, as these concerns exist regardless of whether such bubbles are freely chosen, imposed through classification, or, as is often the case, some mix of the two.

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. For example, a system that determined whether to offer individuals a discount on a purchase based on a seemingly innocuous array of variables being positive ("shops for free weights and men's shirts") would in fact routinely offer discounts to men but not women. To avoid unintentionally encoding such an outcome, one would need to know that men and women arrayed differently along this set of dimensions. Protecting against this sort of discriminatory impact is advanced by data about legally protected statuses, since the ability to both build systems to avoid it and detect systems that encode it turns on statistics.[9] While automated decisionmaking systems "may reduce the impact of biased individuals, they may also normalize the far more massive impacts of system-level biases and blind spots."[10] Rooting out biases and blind spots in big data depends on our ability to constrain, understand, and test the systems that use such data to shape information, experiences, and opportunities. This requires more data.

Exposing the datasets 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 that all result from classifications that shackle users to profiles used to frame their "relevant" experience, are not privacy problems. While targeting, narrowcasting, and segmentation of media and advertising, including political advertising, are fueled by personal data, they don't depend on it. Individuals often create their own bubbles. Merely *allowing* individuals to peel back their bubbles—to view the Web from some-

---

7. Oscar H. Gandy Jr., *Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems*, 12 ETHICS & INFO. TECH. 29, 37-39 (2010).

8. *See generally* ELI PARISER, THE FILTER BUBBLE: HOW THE NEW PERSONALIZED WEB IS CHANGING WHAT WE READ AND HOW WE THINK (2011).

9. Julie Ringelheim, *Processing Data on Racial or Ethnic Origin for Antidiscrimination Policies: How to Reconcile the Promotion of Equality with the Right to Privacy?* 14-15 (Ctr. for Human Rights & Global Justice, Working Paper No. 8/06, 2006) (discussing the use of demographic data to identify disparate impact in "neutral" rules).

10. Gandy Jr., *supra* note 7, at 33.

one else's perspective, devoid of personalization—does not guarantee that they will.[11]

Solutions to these problems are among the hardest to conceptualize, in part because perfecting individual choice may impair other 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 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.[12] 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 sociotechnical system. The law's view of automated decisionmaking systems is schizophrenic, at times viewing automated decisionmaking with suspicion and distrust and at others exalting it as the antidote to the discriminatory urges and intuitions of people.[13] Viewing the problem as one of machine versus man misses the point. The key lies in thinking about how best to manage the risks to the values at stake in a sociotechnical system.[14] 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 about math and machines only second.

Third, lawyers and technologists must focus their attention on the risks of segmentation inherent in classification. There is a broad literature on fairness in social choice theory, game theory, economics, and law that can guide such work.[15] Policy solutions found in other areas include the creation of "standard

---

11. *See* Introna & Nissenbaum, *supra* note 2; Pasquale, *supra* note 2.

12. Recent symposia have begun this process. *E.g.*, Symposium, *Transforming the Regulatory Endeavor*, 26 BERKELEY TECH. L.J. 1315 (2011); *see also* N.Y. Univ. Steinhardt Sch. of Culture, Educ., & Human Dev., *Governing Algorithms: A Conference on Computation, Automation, and Control* (May 16-17, 2013), http://governingalgorithms.org.

13. *See, e.g.*, FED. FIN. INSTS. EXAMINATION COUNCIL, INTERAGENCY FAIR LENDING EXAMINATION PROCEDURES 7-9 (2009).

14. *See, e.g.*, Roger Brownsword, *Lost in Translation: Legality, Regulatory Margins, and Technological Management*, 26 BERKELEY TECH. L.J. 1321 (2011).

15. Among the most relevant are theories of fairness and algorithmic approaches to apportionment. *See, e.g.*, the following books: HERVÉ MOULIN, FAIR DIVISION AND COLLECTIVE WELFARE (2003); JOHN RAWLS, A THEORY OF JUSTICE (1971); JOHN E. ROEMER,

offers"; the use of test files to identify biased outputs based on ostensibly unbiased inputs; required disclosures of systems' 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—policymakers, for example—to have greater influence over, and comfort with, the fairness of classifications. Test files could be used to ensure outcomes comport with this predetermined 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 toward a more complete set of objections.

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, and street corners—where a measure of randomness and unpredictability yields a mix of discoveries and encounters that contribute to a more informed populace. These objections 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,"[16] rather than catering to consumers narrowly focused on private lives, consumption, and infotainment. Equally important, 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

---

EQUALITY OF OPPORTUNITY (1998); JOHN E. ROEMER, THEORIES OF DISTRIBUTIVE JUSTICE (1996); H. PEYTON YOUNG, EQUITY: IN THEORY AND PRACTICE (1995). Roemer's approach to equal opportunity embraces (potentially sensitive) information about the individual over which she has no control—genes, family background, culture, social milieu—explicitly taking these into account, in the form of what he calls a "type," when considering how resources should be allocated.

16. Tanni Haas & Linda Steiner, *Public Journalism: A Reply to Critics*, 7 JOURNALISM 238, 242 (2006).

most important."[17] We want to harness the power of the hive brain to expand our horizons, not trap us in patterns that perpetuate the basest or narrowest versions of ourselves.

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

17. 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), http://culturedigitally.org/2011/10/can-an-algorithm-be-wrong.