Digital Decisions: Policy Tools in Automated Decision-Making

A.R. Lange

ABSTRACT:

Digital technology has empowered new voices, made the world more accessible, and increased the speed of almost every decision we make as businesses, communities, and individuals. Much of this convenience is powered by lines of code that rapidly execute instructions based on rules set by programmers (or, in the case of machine learning, generated from statistical correlations in massive datasets)—otherwise known as algorithms. The technology that drives our automated world is sophisticated and obscure, making it difficult to determine how the decisions made by automated systems might fairly or unfairly, positively or negatively, impact individuals. It is also harder to identify where bias may inadvertently arise. Algorithmically driven outcomes are influenced, but not exclusively determined, by technical and legal limitations. The landscape of algorithmic decision-making is also shaped by policy choices in technology companies and by government agencies. Some automated systems create positive outcomes for individuals, and some threaten a fair society. By looking at a few case studies and drawing out the prevailing policy principle, we can draw conclusions about how to critically approach the existing web of automated decision-making. Before considering these specific examples, we will present a summary of the policy debate around data-driven decisions to give context to the examples raised. Then we will analyze three case studies from diverse industries to determine what policy interventions might be applied more broadly to encourage positive outcomes and prevent the risk of discrimination.

Ali Lange is a policy analyst on the Privacy & Data team at the Center for Democracy and Technology. Her work focuses on empowering users to control their digital presence and includes developing standards for fairness and accountability in algorithms, updating estate law to protect digital accounts from inappropriate access, and researching workplace privacy norms. She has an M.A. in Applied Economics from Johns Hopkins and B.A. degrees in Political Science and English from the University of Chicago.

Jim Dempsey contributed to this paper while he was working at the Center for Democracy & Technology.

INTRODUCTION & BACKGROUND:

In the summer of 2014 Ben Bernanke, former chair of the U.S. Federal Reserve, was denied a mortgage. Since stepping down from his government post earlier in the year, Bernanke had been able to command a reported $250,000 for giving a single speech and signed a book contract estimated to be in the seven figures. Yet when he and his wife sought a mortgage to refinance their house in the District of Columbia, a house whose value according to tax records is dwarfed by his likely income in the next couple years, they were turned down. As the *New York Times* explained, "Ben Bernanke…is as safe a credit risk as one could imagine. But he just changed jobs a few months ago. And in the thoroughly automated world of mortgage finance, having recently changed jobs makes you a steeper credit risk."[1] The numbers were crunched and the decision was made—Mr. Bernanke was denied. Presumably, the former Fed chair found a banker willing to look more closely at his application and reconsider. But how do less well-resourced individuals fare when decisions about credit and other matters of economic consequence are largely arbitrated by the technical systems of financial institutions?

Almost every sector of the economy has been transformed in some way by algorithms. Some have benefited everyone equally by predicting factual outcomes more accurately—producing better weather forecasts, for example. Others empower tools, such as Internet search engines, that are indispensable in the information age. These advancements are not limited to traditionally computer-powered fields; algorithms can read X-rays more accurately than highly trained doctors, at much lower cost.[2] Wall Street fortunes depend on who can write the best trade-executing program. At least one company assesses Hollywood scripts on 100 different points, compares them algorithmically to past box office winners and losers, and predicts success or failure (offering advice to studios on which scenes to drop or new ones to include to boost prospective ratings).[3] For several years now, algorithms have been writing sports stories, taking box scores and converting them into grammatically correct prose virtually indistinguishable from human-authored text. [4] Through online dating and matchmaking sites, millions submit themselves to algorithms hoping to find "the one." Even the government is getting in on the game. Predictive policing technology, powered by algorithms, has spurred a trend in data-driven patrol assignments[5] and parole decisions.[6]

---

[1] Neil Irwin, "Why Can't This Man Refinance?," New York Times, Oct. 3, 2014, p. B1.
[2] Berry de Bruijn et al, "Identifying Wrist Fracture Patients with High Accuracy by Automatic Categorization of X-ray Reports," *Journal of the American Medical Informatics Association 13* (2006), http://jamia.bmj.com/content/13/6/696; Kitty K. Lo et al, "Automatic Classification of Time-Variable X-ray Sources," *The Astrophytical Journal 786* (2014), http://iopscience.iop.org/0004-637X/786/1/20/
[3] Tom Rowley, "The geeks who are directing Hollywood," *The Telegraph*, Jan. 2, 2014, http://www.telegraph.co.uk/culture/film/film-news/10547268/The-geeks-who-are-directing-Hollywood.html. Another algorithm selects hit songs. Christopher Steiner, *Automate This: How algorithms took over our markets, our jobs, and the world* (New York: Portfolio / Penguin, 2012), 77-83.
[4] The practice has spread well beyond sports stories. In March, an algorithm wrote the LA Times story about an earthquake aftershock, drawing data from the USGS Earthquake Notification Service. Gregory Ferenstein, "An Algorithm Wrote The LA Times Story About The City's Earthquake Aftershock Today," *TechCrunch*, Mar. 17, 2014, http://techcrunch.com/2014/03/17/an-algorithm-wrote-the-la-times-story-about-the-citys-earthquake-aftershock-today/.
[5] Heather Kelly, "Police embracing tech that predicts crimes," *CNN*, May 26, 2014, http://www.cnn.com/2012/07/09/tech/innovation/police-tech/.

Technology that helps us make decisions, or makes decisions about us, is inescapable.

Without a doubt, big data and the analytic capabilities that have accompanied it offer huge benefits to consumer-facing businesses. Last year, for example, a study of data brokers by the Federal Trade Commission (FTC) found that risk mitigation products are effective in reducing fraud.[7] The benefits may be especially significant, the FTC concluded, for small businesses and entrepreneurs, which can use data analytics to identify and connect with customers they may not have otherwise been able to reach.

Algorithms have also become standard in making decisions about three fundamental economic opportunities: hiring, insurance, and credit. In these contexts, algorithms assess the risk or benefit posed by an individual to a company's bottom line. On the surface, the relevance of the correlations measured by an algorithm can seem justifiable. Recent job change suggests economic instability and therefore a higher risk of default. Late night driving may indicate frequent barhopping and thus a higher risk of incurring an insurance loss. Since the cost of employee turnover is a financial burden on businesses, applicants whose resumes indicate a series of short-term jobs are less desirable. When it appears that the riskiness or value of an individual can be modeled, mathematically, based on past behavior, algorithms seem a natural fit for these decisions.

However, such uses of algorithms, largely opaque and unregulated, pose questions of fairness and discrimination. Late night driving as a factor in setting auto insurance rates may discriminate against commuters or individuals in low-paid jobs performed at night. Frequent job change may have no correlation to productivity or creativity but rather reflect family circumstances. Indeed, algorithms that select applicants who match the characteristics of the best current employees may be inadvertently replicating past biases, ultimately defeating efforts to attract a more diverse workforce with a wider range of skills and perspectives.

These concerns have been raised at the highest levels of government—the White House commissioned a big data report that concluded, "big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace."[8] If data miners are not careful, sorting individuals by algorithm might create disproportionately adverse results concentrated within historically disadvantaged groups.[9] Current laws on fair credit, equal opportunity, and anti-discrimination may not be adequate to address newer ways of ranking and scoring individuals across a range of contexts. For example, the concept and practicality of redress may be meaningless if an individual does not even know she is being assessed—much less what the criteria are.

---

[6] Jacob Kastrenakes, "Prisons turn to computer algorithms for deciding who to parole," *The Verge*, Oct. 14, 2013, http://www.theverge.com/2013/10/14/4836640/parole-boards-using-computer-systems-determine-recidivism-chances.

[7] "Data Brokers: A Call for Transparency and Accountability," Federal Trade Commission, May 2014, http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

[8] "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

[9] Solon Barocas, and Andrew D. Selbst, "Big Data's Disparate Impact," *SSRN* (September 14, 2014), http://ssrn.com/abstract=2477899.

Algorithms learn based on the training data and human-defined inputs and selection criteria, which means that discrimination can be "baked in" to the process from the beginning. While it is still occasionally uttered,[10] it is naïve to argue that decisions reached algorithmically are unbiased. The computer can only work with the parameters and information it is given and the task of the data miner is in her ability to define target variables and associated characteristics.[11] An algorithm is useful for identifying trends based on statistical correlation and, in the right hands, can sometimes be used to accurately predict a specific outcome. However, it can only predict the future based on the past—or more specifically on whatever data about past events is on hand. Because of this, the results can unintentionally be discriminatory or exacerbate inequality — "garbage in, garbage out" (short hand for "put biased data in, get biased results out") is one of the potential problems of using this technology to make decisions.

That said, hard evidence of algorithmic discrimination is somewhat difficult to come by. Empirical research requires either a fortuitous revelation of potentially widespread disparity (subsequently verified through technical methods), like Dr. Latanya Sweeney's accidental insight that searching her own name prompted ads for an arrest record to be served while searching traditionally white-sounding names did not,[12] or time-consuming forensics that require computer scientists to build profiles of many different consumers and test for differences like the *Wall Street Journal*'s report that Staples varies prices of office products based on the user's proximity to competing stores.[13] Still, researchers and advocates have built considerable work on a constellation of small insights, laying out the potential threats to civil rights and advocating for systems of accountability (if not prudence) on the front end.

SEEKING SOLUTIONS:

Researchers and companies are already developing and implementing algorithms in ways that advance concepts of fairness. Already, for example, the San Francisco-based Evolv, which offers employee selection services, excludes distance from work from its algorithm. Even though their data indicates that workers who live farther away from the work location are more likely to quit, Evolv's ranking does not consider distance from the workplace because it could have an adverse impact on minorities and low-income applicants often stratified in communities distant from areas of economic growth and job opportunities.[14]

---

[10] Cathy O'Neil, "Gillian Tett gets it very wrong on racial profiling," *MathBabe Blog* August 25, 2014, http://mathbabe.org/2014/08/25/gilian-tett-gets-it-very-wrong-on-racial-profiling/. Relevant underlying quote: "After all, as the former CPD computer experts point out, the algorithms in themselves are neutral. "This program had absolutely nothing to do with race… but multi-variable equations," argues Goldstein."
Quentin Hardy, "Using Algorithms to Determine Character," *New York Times*, July 26, 2015. "'Algorithms aren't subjective,' (Jure Leskovec) said. 'Bias comes from people.'"

[11] Presentation, Solon Barocas, "Framing the Conversation," at Federal Trade Commission Workshop, "Big Data: A Tool for Inclusion or Exclusion," Sept. 15, 2014.

[12] "Racism if Poisoning Online Ad Delivery, Says Harvard Professor," *MIT Technology Review*, Feb. 4, 2013, http://www.technologyreview.com/view/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/

[13] Jeremy Singer-Vine et al, "How the Journal Tested Prices and Deals Online," *Wall Street Journal,* Dec. 23, 2012, http://blogs.wsj.com/digits/2012/12/23/how-the-journal-tested-prices-and-deals-online/.

[14] Dustin Volz, "Silicon Valley Thinks it Has the Answer to its Diversity Problem," *National Journal*, Sept. 26, 2014, http://www.nationaljournal.com/next-america/economic-empowerment/can-companies-use-big-data-to-fight-racist-and-sexist-hiring-practices-20140926.

Designed with conscious attention to fairness, algorithmic decision-making systems can avoid at least some discriminatory outcomes and can even expand opportunities.

While some have focused on increasing the transparency of these systems as solution, there are arguments against the effectiveness of transparency. For example, many modern technology companies protect their assets by considering them trade secrets rather than attempting to patent their innovations. This greatly increases the threat of exposing any piece of the system to a public audience. Second, in many cases releasing the criteria that motivate automated decisions creates an opportunity for third parties to game the system. The arms race between spam websites and search engines is a useful illustration of how complicated it can be to be transparent about the equations that determine the order of results.[15] Finally, the temptation to demand complete transparency may be misguided because most regulators, advocates, and interested individuals are not equipped to deal with raw code, or even anything approximating it. As a result, private companies have rejected these unsophisticated demands for transparency.

This has caused advocates, government agencies, and academics to articulate more nuanced principles and techniques to empower individuals affected by automation. A coalition of prominent civil rights organizations,[16] collectively the Civil Rights Roundtable, proclaimed "Civil Rights Principles for the Era of Big Data" which identify a few big-picture priorities describing principles that would help ensure that technology is "designed and used in ways that respect the values of equal opportunity and equal justice."[17] These principles specify that, in order for technology to create economic opportunity for everyone, its creators must, "ensure fairness in automated decisions." While the authors do not describe exactly what they mean by fair, they do propose that technology, "must be judged by its impact on real people, must operate fairly for all communities, and in particular must protect the interests of those that are disadvantaged or that have historically been the subject of discrimination." This foundational update to our understanding of civil rights helps advocates respond to concerns raised by automation and data-based decision-making.

Others have proposed legal or technical solutions. Some have argued existing anti-discrimination laws provide guidance on how to regulate automated systems like online advertising and marketing.[18] Other legal scholars have focused on extending the application

[15] The algorithm Google uses to deliver search results is called Pagerank. It is a patented technology, allowing the company atypical flexibility in revealing how their algorithm works. Even so, revealing too much would allow spammers to manipulate the results, undermining the site's ability to provide results that users want. "How Search Works," Google, https://www.google.com/intl/en_us/insidesearch/howsearchworks/index.html.

[16] Signed by: American Civil Liberties Union, Asian Americans Advancing Justice, Center for Media Justice, ColorOfChange, Common Cause, Free Press, The Leadership Conference on Civil and Human Rights, NAACP, National Council of La Raza, National Hispanic Media Coalition, National Urban League, NOW Foundation, New America Foundation's Open Technology Institute, and Public Knowledge.

[17] Letter from Civil Rights Organizations, "Civil Rights Principles for the Era of Big Data," 2014, http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html.

[18] Peter Swire, "Lessons From Fair Lending Law for Fair Marketing and Big Data," Submitted to the Federal Trade Commission, Sept. 11, 2014, http://www.futureofprivacy.org/wp-content/uploads/FairMarketingLessons_WhitePaperFTC.pdf. "Fair lending law provides decades of lessons from jurisprudence, regulatory guidance, and industry initiatives that could help guide next steps for both industry and the FTC to evaluate marketing data and ensure advertising is done in ways that benefit consumers" (11).

of basic legal principles, like due process,[19] to address new challenges posed by digital technology. Academic and industry researchers have focused their efforts on testing technical solutions or restrictions. Computer science researchers have proposed technical solutions to ensure fairness, ranging from statistical parity[20] to auditing mechanisms designed to check against the disparate impact test the government uses in federal housing cases.[21] Social scientists have described "Big Data's Disparate Impact,"[22] showing the ways that data mining techniques can create disparate outcomes. Focusing more broadly, the Federal Trade Commission hosted a workshop to bring together stakeholders to debate the merits of big data and the potential role for government in protecting underserved consumers from harm.[23]

Despite the extent of discussion, debate, research, and negotiation—all guided by a general consensus that this is an important problem—it is not clear what a large-scale effort to prevent discrimination in automated decision-making would look like. One barrier is the scope and diversity of contexts in which this technology is deployed. Rather than creating rules and restrictions, it seems more productive to guide engineers, designers, and data scientists with basic principles. The Civil Rights Roundtable has taken the first step by laying a foundation of expectation, but these are not specific enough to be deployed by technologists. But if we incorporate the perspective of the advocacy community, we can then draw from existing examples of changes made by private companies, both as a result of external pressure and internal values, and draw conclusions about useful interventions.

CASE STUDIES:

The White House Big Data Report called for a "national conversation on big data, discrimination, and civil liberties."[24] But we do not have to wait for a national dialogue—we can examine existing systems to understand how internal policy choices factor into algorithmic outcomes. This section will analyze the prevailing policy incentives in three case studies from diverse industries, with accompanying recommended interventions that might mitigate bias and expand positive outcomes.

The effect of automated decision-making systems on individuals is influenced by internal policy choices as well as the external regulatory landscape. The prevailing principle in each

[19] Kate Crawford, and Jason Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," *Boston College Law Review 55* (Jan. 29, 2014), http://lawdigitalcommons.bc.edu/bclr/vol55/iss1/4/. "(P)rocedural data due process would regulate the fairness of Big Data's analytical processes with regard to how they use personal data (or metadata derived from or associated with personal data) in any adjudicative process, including processes whereby Big Data is being used to deter- mine attributes or categories for an individual" (109).
[20] Cynthia Dwork, Moritz Hardt et al, "Fairness Through Awareness," arXiv:1104.3913v2 [cs.CC] (Nov. 2011), http://www.cs.toronto.edu/~zemel/documents/fairAwareItcs2012.pdf.
[21] Sorelle Friedler, Carlos Scheidegger and Suresh Venkatasubramanian, "Certifying and Removing Disparate Impact," arXiv:1412l3756 [stat.ML] (Dec. 2014), http://arxiv.org/abs/1412.3756.
[22] Solon Barocas, and Andrew D.Selbst, "Big Data's Disparate Impact," *California Law Review 104* (February 13, 2015), http://ssrn.com/abstract=2477899.
[23] Workshop, "Big Data: A Tool for Inclusion or Exclusion," Federal Trade Commission, Sept. 15, 2014, https://www.ftc.gov/news-events/events-calendar/2014/09/big-data-tool-inclusion-or-exclusion.
[24] "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. Page 59.

of these cases is illustrative of the role that policy decisions play in shaping outcomes for individuals. The positive outcomes for users and consumers highlight the importance of engaging private industry directly in addressing the risk of automated discrimination. Examining the principles behind decisions by private companies demonstrates that there are multiple ways to ensure that individuals are treated fairly and with respect by automated systems.

I. Give Meaningful Access to Data & Align Incentives for Consumers: Credit and the Financial System

Approval for credit is one of the oldest data-driven decisions, yet as we saw with Mr. Bernanke's example, it is still an imperfect science. At least, as a result of the policies regulating the credit industry we can draw a direct line between the data and the result: Bernanke's recent job change had caused a lower credit score, resulting in the denial of the loan. Access to credit data, and an explanation when credit is denied or revoked, was a victory of a legislative battle in the 1970s that brought about the Fair Credit Reporting Act (FCRA). This law established foundational rights for consumers in their interactions with credit reporting bureaus in the face of an increasing data collection and aggregation. Prior to this legislation, credit-reporting bureaus were shadowy agencies whose files were filled with records of drinking, marital discord, and adultery.[25] Providing individuals with a right to access their report and correct it was revolutionary, and access remains a pillar of consumer rights advocacy.[26] However, in a world where we generate data far faster than we can analyze or review it,[27] providing access to raw data is no longer sufficient consumer protection.

Perhaps one of the most positive outcomes of automation is fraud detection technology. While credit card companies do offer near-real-time access to a customer's transaction data, constantly monitoring it for potential fraud would be hugely burdensome for individuals. Using algorithms to monitor credit card accounts for unauthorized charges saves people the trouble of vigilantly monitoring their accounts while reducing the cost of fraud to the overall financial system.

This innovative technology was not developed solely out of dedication to the integrity of the system, however. It was encouraged by the policy decision to shift the financial burden for fraud onto institutions rather than individual account holders. Shifting the majority of the burden of fraud against individual financial accounts to the institution was part of the Electronic Fund Transfer Act.[28] In addition to significantly limiting the financial liability a consumer may face, this regulation also places the burden of proof on the institution. The administrative burden of collecting the evidence necessary to prove that the consumer, in fact, authorized an electronic transfer encouraged financial institutions to avoid the task outright. Instead, financial institutions moved to take advantage of the huge amount of data

[25] 115 Cong Rec 2410 (Jan 31, 1969).
[26] Fair Information Practice Principles, www.nist.gov/nstic/NSTIC-FIPPs.pdf.
[27] "Are you ready for 2016: The Year of the Zettabyte," March 23, 2013. http://dailyinfographic.com/2016-the-year-of-the-zettabyte-infographic. According to Cisco Systems, the world will cross the threshold into one Zettabyte of data in 2016. One Zettabyte, "would be able to store just over two billion years of music."
[28] Pub. Law 95-630, 92 Stat. 3641 (1978).

created by each individual account holder to develop a profile of fraud and automate an early-detection system.[29] Automating the process of fraud detection, and erring on the side of false-positives, ensured that the losses to the institution (either as a direct result of fraud or the administrative burden of proving otherwise) were minimized.

In this case, where accuracy is a relatively straightforward and high-stakes game, holding financial institutions responsible for safeguarding the funds they have been entrusted with is a logical consumer protection measure. This expectation can be expanded to other institutions with similar power over people's lives; when the decision affects an individual's bottom line, the burden of accuracy should fall to the institution collecting and using the data, rather than to the individuals effected by the results.

This example demonstrates a general example of how automated systems can create gains for individuals—not only can one see the data itself, but also the institution shares the benefits of the analytics. While the letter of the law did not mandate an automated system, the incentive structure created by it encouraged its innovation. By comparison, the harm of inaccurate credit reports (and the missed opportunities for efficiency in the overall system) provides an example of an incentive structure that discourages ideal outcomes. There is an opportunity here for a policy intervention that re-balances power between individuals and institutions collecting data about them by raising the expectations of access and shared gains of analytics.

The principle of access and incentivizing innovation can be applied beyond the financial system to help mitigate civil rights violations on the ground. While individuals shouldn't be responsible for monitoring the integrity of a system, they should be able to report anomalous or confusing results. Allowing an individual to view and assess data collected about them creates a backstop against an incorrect or unfair decision. Creating a feedback loop for users to share information gives companies information on whether the decisions of an automated system demonstrate patterns of unfairness, bias, or discrimination. The majority of automated decisions do not require review, but installing a mechanism for aggregate observation is imperative to catching and remedying unintended discrimination. Embracing access as a principle encourages companies to create technology for users to report their experiences once an automated system is publicly deployed or widely used.

II. Explain the Logic: The Power of Page Rank

Almost 68 percent of Internet searches in the United States are conducted by Google's engine.[30] Thus Google's internal policies and decision-making equations play a tremendous role in determining the public's access to information online. The internal policies of tech giants like Google materially inform our experience of the Internet and the world. For example, Google's search algorithm reflects, and at times distorts, human proclivities, as in the highly publicized example of the wife of a German politician whose name prompted

[29] Philip Chan and Wei Fan, "Distributed Data Mining in Credit Card Fraud Detection," *IEEE Intelligent Systems* (Nov/Dec. 1999), http://cs.fit.edu/~pkc/papers/ieee-is99.pdf.
[30] Press Release, "ComScore Releases March 2014 U.S. Search Engine Rankings," ComScore, Apr. 15, 2014, http://www.comscore.com/Insights/Press-Releases/2014/4/comScore-Releases-March-2014-U.S.-Search-Engine-Rankings. The reported figure is 67.5 percent.

autocomplete suggestions of "prostitute" and "escort." When she sued Google, the company's response was that the results were algorithmically determined as a byproduct of the aggregate searches being conducted.[31] (Perhaps ironically, Ms. Wulff's name is now also associated with a significant amount of academic scholarship and media attention on the role Google plays in reputation management and access to information.) Ultimately, the company changed its policy to remove autocomplete suggestions for *all* names, except those of public figures.[32] Still, the criteria for what constitutes a public figure are not quite clear nor is it clear when or why Google's search results are modified in response to how famous someone is. The company has made similar adjustments to its algorithm in response to another media frenzy around sites that blackmailed individuals by posting mug shots and only removing them for huge fee,[33] and it is currently in the process of deciding how to manage requests to takedown revenge porn.[34]

More generally, these examples demonstrate that technology companies make dynamic and fast-paced policy decisions, reflecting the nature of their products. These decisions affect a huge number of people, as they determine both what information we have access to and what information about us is promoted as the "most relevant." But the proprietary nature of the technology industry complicates the call for transparency.

How are these algorithmic policy decisions made? The rubric for whether a company like Google decides to change a policy, or explain the logic behind a particular automated result, appears to exist on a plane between embracing societal values and the pressures of being a public-facing company. It may be the case that the decisions are determined in part by whether a public explanation of the inferences behind a result would be considered generally logical, or if it might embarrass the company by being tone-deaf or fundamentally biased.

A company determining how to apply this sort of common-sense logic might draw useful guidance from the idea of 'inferential privacy,'[35] a term coined by Solon Barocas that looks to cultural norms as a guideline. Specifically, he argues that a company should not infer individual characteristics from data if it would not be asked in person. For example, if you would not ask a teenage girl if she is pregnant face-to-face, you should not try to deduce that from her browsing history and send her relevant advertisements.[36] This argument responds to the protestation that it is difficult to know where the line is with marketing—sometimes a

---

[31] Meena Hart Duerson, "Bettina Wulff Sues Google: She was never an 'escort' or 'prostitute!' *New York Daily News*, Sept. 11, 2012, http://www.nydailynews.com/news/world/bettina-wulff-sues-google-escort-prostitute-article-1.1156819.

[32] There are other situations under which Google's search results are modified in response to how famous someone is—perhaps looking through the decisions made there could shine light on the criteria. Stewart Baker, "Does Google Think You're Famous," *The Volokh Conspiracy*, published in *The Washington Post*, Sept. 1, 2014, http://www.washingtonpost.com/news/volokh-conspiracy/wp/2014/09/01/does-google-think-youre-famous/.

[33] David Segal, "Mugged by a Mug Shot Online," New York Times, Oct. 5, 2013, http://www.nytimes.com/2013/10/06/business/mugged-by-a-mug-shot-online.html?_r=0. "If it acted, Google could do what no legislator could — demote mug-shot sites and thus reduce, if not eliminate, their power to stigmatize."

[34] Joanna Walters, "Google to exclude 'revenge porn' from internet searches," The Guardian, June 21, 2015, http://www.theguardian.com/technology/2015/jun/20/google-excludes-revenge-porn-internet-searches.

[35] Solon Barocas, "Leaps and Bounds: Toward a Normative Theory of Inferential Privacy" (Forthcoming).

[36] Kashmir Hill, "How Target Figured Out a Teen Girl was Pregnant Before Her Father Did," *Forbes*, Feb. 16, 2012, http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

pregnant woman would appreciate diaper coupons or a warning about a dangerous medication interaction. Creating a rubric focused around cultural norms is a moving target, but one that is unlikely to draw public ire, assuming it stays current. A norms-based philosophy works for a company that is not looking to execute a moral decision through its automated system but rather only to stay out of the news.

Decision makers need to be explicit about the logic of their decision and attempts to provide a standard for doing so that balances the need for transparency with the reputational interests of the entity. Ultimately, companies caught in the spotlight for violations of anti-discrimination protections often find themselves with two choices: address the underlying concern or explain why you will not. While the second option could seem like a dead end, the example of Google's changed response to Ms. Wulff's case demonstrates that this strategy can also lead to productive outcomes. The company's explanation of how they saw themselves in the information economy (that its duties lie strictly in providing information that reflects the zeitgeist) provided insight into the culture and policies of an otherwise obscure process.

Technology companies do respond to concerns raised by the media, advocates, academics, and individuals, indicating that their logic is as dynamic as the algorithms themselves. The policies set internally at these companies have a huge and pervasive impact on the constraints imposed on algorithmic decision-making and, therefore, on the experience of the individual as she travels around the Internet. Therefore, advocacy that focuses on influencing internal policies (tempered by a reasonable understanding of the incentive structures faced by companies) can create more egalitarian outcomes relatively quickly and across the board. Ms. Wulff's example demonstrates that direct-to-company advocacy is a productive way to prevent biased outcomes perpetuated by automated systems.

III. Interrogate the Data: Failure of Predictive Policing

Predictive policing is often an example of a failure to make inferences from correlations in data ethically. The incentive structures around automation and data by government entities are less well defined. In general, law enforcement is tasked with reducing violence and crimes to zero, and the tools and techniques to achieve this mandate are accordingly vast. However, incentives to protect other values are less direct and it can be difficult for traditional protections such as the Fourth Amendment of the U.S. Constitution to stay apace of new technology. Local and state law enforcement offices have begun to use sophisticated data analytics and algorithms in an attempt to be more effective, including tools that claim to provide predictions of where crimes will happen (or who will commit them). Other tools are applied to the parole process to predict the likelihood of re-offending. The pressure to lower crime rates, along with a lack of expertise in data and machine learning, has resulted in police forces with potentially biased inferences that lack oversight or critical consideration before they are deployed.

A notorious example of this problem is Chicago's Custom Notification Program, more

commonly referred to as the "heat list",[37] which is an algorithmically determined list of names that purports to predict who is likely to commit a crime. The underlying theory is based on research by a sociologist who showed that revenge is a primary motivator of violent crime. He argues that by tracking social connections among offenders, we can anticipate who is likely to respond violently to a crime committed against an associate. The execution of his theory relies on tracking offenders who are arrested together and attempting an early intervention if the system flags someone as likely to commit a crime. It is controversial at best, and at worst latently racist. Without knowing what criteria the program considers, there is no way to assure that the inferences it is making are fair and unbiased. It is likely that the results of the computer program reflect a pervasive unfairness in the real world—building a program that relies on social connections in one of the most racially divided cities in the country is likely to result in racially skewed results and while using a predictive model based on historic data has the obvious flaw of potentially perpetuating whatever bias went into the original dataset.

In the case of government use of automated decision-making, transparency is the right policy response. It is difficult to make an argument that the technology used to assist police in predicting crime should be proprietary. The claims that transparency of the criteria allows people to game the system, for example, absolutely do not apply to a system where the goal is to reduce the underlying behavior. For example, assuming the criteria for crime prediction are at all related to committing crimes, gaming the system would mean not committing crimes. But more importantly, there is something fundamentally awry with the idea that a decision about police actions can be motivated by a black box with no oversight or public scrutiny. Despite the fact that these tools are often used as part of a public-private partnership (as in the case of the heat list), they must come with a condition of transparency or the risks for civil rights violations are simply too high.

No form of decision-making is perfectly accurate or free of bias—humans have always misjudged others, and using algorithms as a decision-making tool has not changed that fundamental reality. Without knowing the ground truth (i.e. a factual measure of reality) it is nearly impossible to know the true error rate of a system that makes decisions about individuals. However, as part of the design for an automated system, someone has to make decisions about what kinds of outcomes are acceptable or expected and train the technology to detect and elevate these trends. Unfortunately, some technology evangelists still believe that using math means a decision is fair. Everyone engaging with an automated decision-making system, from the engineer to the product development team to the individual, needs to be critical about the origination of the data behind the model.

CONCLUSIONS:

The concerns raised about the secrecy of data-driven decision-making are growing in urgency and volume. The promise of civil rights is that individuals are treated fairly, irrespective of race, class, sex, or disability. We have laws that protect individuals from discrimination on these traits and, as the White House Big Data Report concluded, "big data

[37] Jay Stanley, Chicago Police "Heat List" Renews Old Fears About Government Flagging and Tagging," *ACLU*, Feb. 25, 2014, https://www.aclu.org/blog/chicago-police-heat-list-renews-old-fears-about-government-flagging-and-tagging.

analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace."[38] Stakeholders agree that civil rights laws should be honored by new technologies, but the diversity and dynamism of the technology market complicate the question of how to intervene. A call for transparency is rooted in historic advocacy victories, but transparency is a blunt instrument and this is a nuanced problem.

Rather than focusing on the solutions that leveled the playing field in the 1970s and 1990s, we should consider cutting-edge policy solutions built on the premise that the individual should share in the benefits of technical sophistication. This requires that companies start with the assumption that individuals should be treated as partners in, rather than inputs for, a data-driven process. While not exactly the same as the guiding philosophy of the civil rights movement (which also argues that historically disadvantaged groups of people should be protected), this philosophy lays the foundation for embedding fairness and respect in automated technology.

In particular, the following policy interventions have had been successful for a specific use case and they may merit consideration as solutions going forward:

1. *Give Meaningful Access to Data & Align Incentives for Consumers:*

    - Engage users directly in a data driven decision to create value for individuals.
    - Align incentives to encourage companies to innovate automation that protects individuals by placing appropriate burdens on companies.
    - Encourage companies to create a direct feedback loop for users or customers to report potentially unfair or biased decisions.

2. *Explain the Logic of the Decision:*

    - Give insight into technical or policy limitations and create public dialogue.

3. *Interrogate the Data:*

    - The context of the underlying data limits its usefulness and should limit its role in the decision.

These are not the only solutions to discrimination in data-driven systems. The Center for Democracy & Technology is working with technology companies, the civil rights community, and academics to identify other fundamental principles that can help mitigate the risk of discrimination. The variety of solutions being proposed is novel because the problem is diffuse and often unintended. However, looking to existing policy outcomes and decisions demonstrates that a few principles could, if embedded in the design and development process, mitigate the risk for discrimination and other harms from bias and automation. The first step is to identify what questions we need to ask ourselves as we create a process and ensure that we are critically engaging with the system. Additionally, new technology creates

[38] "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

opportunity for innovation and economic gains for both individuals and private companies. By building on a foundation of civil rights principles, we ensure that everyone shares in the spoils equally and that technical innovation continues to propel us forward into the future without undermining the advocacy victories of the past.