Table of Contents

Foreword .................................................................................................................................................. 3
Introduction ........................................................................................................................................... 5
Chapter 1: Financial Inclusion ................................................................................................................ 6
As Insurers Embrace Big Data, Fewer Risks Are Shared ................................................................. 6
Data Brokers Enable Targeting of Financially Vulnerable Communities ........................................... 8
Furthering Financial Inclusion with "Alternative Data" ................................................................... 10
Chapter 2: Jobs ....................................................................................................................................... 12
E-Verify: The Disparate Impact of Automated Matching Programs .................................................... 12
Hiring Algorithms May Put Jobs Out of Reach ...................................................................................... 15
In Online Searches, Big Data Systems Reproduce Racial Bias ............................................................. 16
Chapter 3: Criminal Justice .................................................................................................................. 18
Predictive Policing: From Neighborhoods to Individuals ................................................................... 18
Secrecy Is a Barrier to Responsible Use of Police Technologies ......................................................... 20
Body-Worn Cameras Can Promote Police Accountability ................................................................... 22
Chapter 4: Government Data Collection and Use ................................................................................. 24
Dragnet Surveillance Short-Circuits Court Review in Drug Cases ....................................................... 24
The Temptation to Abuse Government Databases ............................................................................. 25
The Census: Big Data for Civic Welfare ............................................................................................... 26
Acknowledgements ............................................................................................................................... 28
Introduction

The key decisions that shape people’s lives—decisions about jobs, healthcare, housing, education, criminal justice and other key areas—are, more and more often, being made automatically by computers. As a result, a growing number of important conversations about civil rights, which focus on how these decisions are made, are also becoming discussions about how computer systems work.

Earlier this year, a path-breaking coalition of major civil rights and media justice organizations released the Civil Rights Principles for the Era of Big Data, highlighting how the growing use of digital surveillance, predictive analytics, and automated decision-making impacts core civil rights concerns. We served as technical advisors to that coalition.

After the release of the Principles, there was an outpouring of interest from policymakers, community advocates, corporate leaders and the public. People want to know more about the concrete examples that motivate this work. How and where, exactly, does big data become a civil rights issue? This report begins to answer that question, highlighting key instances where big data and civil rights intersect. We hope it will serve as a valuable resource to everyone involved in this important, emerging conversation.

— David Robinson, Harlan Yu, and Aaron Rieke, Upturn (formerly Robinson + Yu)
Data Brokers Enable Targeting of Financially Vulnerable Communities

Both the Federal Trade Commission and the Senate Commerce Committee recently released significant research reports on the data broker industry, which collects enormous volumes of information on hundreds of millions of Americans. The reports detail how these largely-unregulated companies enable precision-marketing of consumer products to financially vulnerable individuals. The Senate report further warned that the data sold by some brokers is “likely to appeal to companies that sell high cost loans and other financially risky products,” and the FTC observed that many would find it “disconcerting,” to know that products can easily be targeted at disadvantaged people. [5]

The Senate report identified marketing lists with titles like “Rural and Barely Making It,” “Ethnic Second-City Strugglers,” ‘Retiring on Empty: Singles,’ ‘Tough Start: Young Single Parents,’ and ‘Credit Crunched: City Families.’” [6] The Commission's report also highlighted segments focused on minority communities and low-income individuals, including a one called the “Urban Scramble.” [7] It also observed that data brokers sell “Assimilation Codes,” indicating a person’s degree of familiarity with the English language. [8] Much of the negative publicity these marketing lists have received stems from their evocative titles—but the fundamental issue runs deeper: the lists enable marketers to identify vulnerable consumers with ease.

Of course, targeted marketing has a place in connecting all communities with the products and services most attractive to them—including for poor consumers, people of color, and people who speak different languages. But precision targeting of vulnerable groups also carries a risk of harm.

Modern data brokerage is an evolution of an old practice. Businesses have a long history of collecting data to help them target or acquire new customers. However, information technology has facilitated a rapid increase in both the volume and availability of data about individuals. Companies are now able to collect and store far more than would have been thought possible in decades past.

“Data broker” is a broad label used to describe the companies that buy, sell, or analyze consumer information. These firms offer marketing services, fraud prevention, risk assessment, data consolidation, or just resell data to other data brokers. There is no comprehensive list of companies that fall under this umbrella. [9]

Data brokers vacuum up data from wherever they can, including from public records, social media sites, online tracking, and retail loyalty card programs. Using these data, brokers build “modeled” profiles about individuals, which include inferences and predictions about them. For example, a broker might infer marital status from the prefix “Mrs.” or wealth based on an individual’s neighborhood. These profiles are often sold in the form of “segments” (or marketing lists) which are priced and sold by the thousands.
There are few laws governing the data brokerage industry, even though many of its practices can resemble the type of consumer scoring that is regulated in other contexts. The Government Accountability Office explained that “consumers generally do not have the right to control what personal information is collected, maintained, used, and shared about them—even where such information concerns personal or sensitive matters about an individual’s physical and mental health.” Similarly, federal law gives consumers the right to correct errors in their credit histories, but no similar right exists with respect to the profiles held by data brokers. The data brokerage industry has been repeatedly criticized for its lack of transparency, and the FTC recently unanimously renewed its call for Congress to enact legislation and empower individuals by allowing them access to information held by data brokers.

This unregulated landscape is a challenge to social justice groups who are mindful of a history of predatory marketing and lending toward vulnerable groups. Data brokers can enable discriminatory targeting based on sensitive information like financial situation, health indicators, or other signs of vulnerability.
Hiring Algorithms May Put Jobs Out of Reach

Many retailers, call centers, and other employers of entry-level service staff have begun using machine learning systems to evaluate job applicants. Analyzing numerous factors for thousands of employees, specialized technology firms develop online questionnaires that surface the factors most predictive of success for each employer and job.

Some firms have found that people with shorter commutes tend to make better hires, because they are statistically likely to stay in the job longer. This insight may be particularly important for service sector employers, whose hiring is increasingly automated, and for whom turnover is a major concern. According to a 2012 Wall Street Journal report, a hiring analytics firm called Kenexa (now owned by IBM) “asks applicants for call-center and fast-food jobs to describe their commute by picking options ranging from ‘less than 10 minutes’ to ‘more than 45 minutes.’ The longer the commute, the lower their recommendation score for these jobs, says Jeff Weekley, who oversees the assessments.”

The same story also notes that how reliable a person’s transportation is (i.e., whether they depend on public transportation) and how long they have lived at their current address may also be considered.

A second firm that applies big data to the hiring process, Evolv, has reportedly made a different choice. As the Atlantic Monthly reported:

There are some data that Evolv simply won’t use, out of a concern that the information might lead to systematic bias against whole classes of people. The distance an employee lives from work, for instance, is never factored into the score given each applicant, although it is reported to some clients. That’s because different neighborhoods and towns can have different racial profiles, which means that scoring distance from work could violate equal-employment-opportunity standards. 

A hiring preference against workers who live far away may be accurate—they may really average shorter tenure in the job—but is it fair? Such a preference punishes people for living far from where the jobs are, and can particularly hurt those living in economically disadvantaged areas, who are disproportionately people of color. Such practices make it even harder for people in disadvantaged communities to work their way out of poverty.
In Online Searches, Big Data Systems Reproduce Racial Bias

Digital indicators of race, religion, or sexual preference can easily be observed or inferred online. In some ways, these indicators are just like those an employer might pick up when scanning a person's resume. However, a recent study has revealed that these indicators can foster “discriminatory outcomes or giv[ing] preference to members of one group over another” when combined with complex big data systems.

Latanya Sweeney, a computer science professor at Harvard who recently served as Chief Technologist at the Federal Trade Commission, described how Google ads discriminate based on the name of the person searched. When searching for her own name on Google, Dr. Sweeney noticed ads referencing arrest records. This prompted her to design a study to learn whether searches for white-identifying names prompted the same sorts of ads as searches for black-identifying names did. She found that a greater percentage ads with “arrest” in their text appeared for black-identifying names than for white-identifying names, to an extent that could not plausibly be explained by chance. She concluded that “[t]here is discrimination in delivery of these ads.”

This happens because Google's software automatically learns which ad combinations are most effective (and most profitable) by tracking how often users click on each ad. These user behaviors, in aggregate, reflect the biases that currently exist across society. Instantcheckmate.com, a leading company that sells arrest records, denied that it has ever tried to connect a name with race. But it would not necessarily have to for this outcome to occur.

Ads that are more often clicked on automatically receive a higher “quality score”—and are more often displayed—in Google's system. Google and InstantCheckmate may automatically find themselves reinforcing the racial biases that their audience's click patterns reflect. Dr. Sweeney explains: “If Google's Adsense service learns which ad combinations are more effective, it would first serve the arrest-related ads to all names at random. But this would change” as the algorithm automatically changed in response to a pattern, where “click-throughs are more likely when these ads are served against a black-identifying name.”

These sorts of structural discrimination issues are particularly troubling as employers—and others in positions of power and responsibility—increasingly consult the Internet when making the decisions that shape people's lives. Although potential employees have some legal protections today, it would be difficult for a job applicant harmed by the subliminal effects of biased ads to trace such harm to its cause. A quick glance (or many such glances) by a hiring professional are likely to go unnoticed. The same concerns may arise in situations involving promotions, special awards, or other forms of professional advancement, or in different settings such as the search for a roommate.

Lawyers do caution employers to tread carefully online. “I advise employers that it's not a good idea to use social media as a screening tool,” says James McDonald, a specialist in employment law. “[Employers] need to control the information,” he says, but the ease of a Google search may be hard
to resist. “By and large, employers avoid asking questions about these traits in interviews. But now technology makes it easier to find that information,” observes Prof. Alessandro Acquisti of Carnegie Mellon University. [51]

Dr. Sweeney’s research shows that racism can be perpetuated inadvertently by complex online systems, even when the companies that create these systems do not intend to discriminate.
Chapter 3: Criminal Justice

Predictive Policing: From Neighborhoods to Individuals

In February 2014, the Chicago Police Department (CPD) made national headlines for sending its officers to make personal visits to residents considered most likely to be involved in a violent crime. The selected individuals were not necessarily under investigation, but had histories that implied that they were among the city's residents most likely to be either a victim or perpetrator of violence.

The officers' visits were guided in part by a computer-generated “Heat List”: the result of an algorithm that attempts to predict involvement in violent crime. City officials have described some of the inputs used in this calculation—it includes some types of arrest records, for example—but there is no public, comprehensive description of the algorithm's input.

The visits were part of a new “Custom Notification Program,” which sends police (or sometimes mails letters) to peoples' homes to offer social services and a tailored warning. For example, officers might offer information about a job training program or inform a person that federal law provides heightened sentences for people with certain prior felonies. The city reports that the contents of a notification letter are based on an analysis of “prior arrests, impact of known associates, and potential sentencing outcomes for future criminal acts.” Although some of these visits have been poorly received, the department argues that the outreach efforts may already have deterred crime. Mayor Emanuel recently claimed that, of the 60 interventions that have already taken place, “none of the notified individuals have been involved in any new felony arrests.”

The Heat List is a rank-order list of people judged most likely to be involved in a violent crime, and is among the factors used to single people out for these new notifications. The CPD reports that the heat list is “based on empirical data compared with known associates of the identified person.” However, little is known about what factors put people on the heat list, and a FOIA request to see the names on the list was denied on the grounds that the information could “endanger the life or physical safety of law enforcement personnel or [some] other person.” Media outlets have reported that various types of data are used to generate the list, including arrests, warrants, parole status, weapons and drug-related charges, acquaintances’ records, having been a victim of a shooting or having known a victim, prison records, open court cases, and victims' social networks. The program's designer, Illinois Institute of Technology (IIT) Professor Miles Wernick, has denied that the “algorithm uses 'any racial, neighborhood, or other such information' in compiling the list.”

Cities across the country are expanding their use of data in law enforcement. The most common applications of predictive technology are to assist in parole board decisions and to create heat maps of the most likely locations of future criminal activity in order to more effectively distribute...
police manpower. Such systems have proven highly effective in reducing crime, but they may also create an echo chamber effect as crimes in heavily policed areas are more likely to be detected than the same offenses committed elsewhere. This effect may lead to statistics that overstate the concentration of crime, which can in turn bias allocations of future resources.

Chicago’s experiment is one of several of a new type, in which police departments move beyond traditional geographic “crime mapping” to instead map the relationships among city residents. Specifically, identifying individuals for tailored intervention is the trend most likely to expand in the future of predictive policing—raising important questions on how to ensure justice continues to be protected through machine systems. Other districts are already working with academics to develop similarly styled programs, including one in Maryland that aims to “predict which of the families known to social services are likely to inflict the worst abuses on their children.”[64] In projects like these, automated predictions of future bad behavior may arise—and may be acted upon—even without direct evidence of wrongdoing. Such systems will sometimes make inaccurate predictions, and when they do, their mistakes may create unjustified guilt-by-association, which has historically been anathema to our justice system.

Even as they expand their efforts to collect data, city governments often do not have the academic resources to analyze the vast amounts of data they are aggregating. They are often partnering with private or academic institutions to assist in the process. In Chicago, the city is working with the MacArthur-backed Crime Lab to analyze the effectiveness of various programs, including things like “Becoming A Man,” a program that focuses on violence prevention among at-risk youth.[65] These partnerships allow the city to expand the ways it uses the data it collects, and may unlock significant benefits (by, for example, demonstrating the effectiveness of non-punitive crime reduction programs). At the same time, the private actors conducting these and other analyses should be held to at least the same standards of accountability and transparency that would apply if the city were analyzing its data internally.