# Algorithmic Bias

**Joni R. Jackson**
**Chicago State University**

*Our lives are being quantified by algorithms that are hidden from view. This paper explores the various ways in which algorithms fuel biased decision making. Algorithms are used to profile and predict people's behavior, presumably in a neutral manner. However, algorithms may be tainted by our biases in covert and subtle ways that can have a significant impact on people's lives.*

## INTRODUCTION

The amount of data created by and about us has grown exponentially. In 2016 alone, we produced as much data as we had created since our existence. The amount of data and the growth in data is unprecedented and expected to continue to grow; by 2028, it is estimated that there will be 150 billion sensors measuring all that we do (Helbing, Frey, Gigerenzer, Hafen, Hagner, Hofstetter, van den Hoven, Zicari, & Zwitter, 2017). In addition to the passive collection of data from sensors, there is the data that we willing volunteer (e.g., via social media) and the data that is inferred by combining existing data to create new data and insights (Gandy, 2016).

The growth of data and the desire for better decision-making has increased efforts to mine this growing repository of data. This data allows us to paint a clearer picture of who people are and, like a crystal ball, to predict their future behaviors. We can use data to search for patterns and correlations that allow us to make predictions about people's future behavior. Even if data is missing, it is possible to impute that which is missing by looking at patterns among people with similar characteristics (e.g., women, Millennials, etc.) or to look for proxy variables (race or ethnicity) assumed to be predictive of behaviors (Williams, Brooks, & Shmargad, 2018).

The use of data to predict outcomes and inform decisions occurs across both public and private spheres, and ranges from decisions as diverse as those related to consumer marketing or public policy. At the center of the data mining efforts are algorithmic models which can combine data in novel ways to create new meaning or to infer meaning - very much like people do - from missing data. Algorithms search for patterns in data. By using past data and inferring meaning from missing data, these models can learn patterns that correlate with social variables (e.g., race or gender) and produce discriminatory results, whether intended or not (Greenwald, 2017; Williams et al., 2018).

Algorithms use data to create and infer meaning; if the meaning that we glean from algorithms is consistent with our experience, then the algorithms may confirm what we already believe (Spence, 1973; Varshney & Varshney, 2017). This confirmation of our beliefs may then be fed back into the model, providing information suggesting that the belief is accurate. As the models learn based on these signals

this may serve to perpetuate biases of which we are not aware. What makes these biases difficult to detect is the assumption that algorithmic (mathematical) models are objective.

This paper will review some of the recent developments in how algorithms are deployed in data capture and analysis and will explore how algorithms impact people's lives in covert and subtle ways that can have a significant impact on their well being.

## ALGORITHMIC MODELS ("ALGORITHMS")

A model is an "abstract representation of some process … that takes what we know and and uses it to predict [behaviors] in various situations" (O'Neil, 2016). Algorithms reduce decision making to a number. This reliance on numbers suggests an objective, unbiased approach to decision making based on the assumption that numbers, unlike people, do not lie. While algorithms are assumed to be neutral, their neutrality is difficult to discern because algorithms are opaque, often described as impenetrable "black boxes" (Bruckner, 2018; O'Neil, 2016; Rieland, 2018). Our biases are embedded in these algorithms, often so deeply as to render our biases virtually invisible (Aguirre, 2000; Coates, 2011; O'Neil, 2016). This invisibility leads to the presumption that algorithms are neutral. However, algorithms often reflect the very biases they are presumed to ignore.

Moreover, the people who build the algorithms often are not diverse. Overwhelmingly White and male, their biases, values and assumptions tend to shape the design of algorithms (Karppinen & Oinas-Kukkonen, 2013; Rayome, 2018; Yetim, 2011). Many voice assistants (e.g., Alexa and Echo) are female (most offer male options now). The selection of female voices reflects (and probably reinforces) stereotypes of women as caregivers and helpers. Professor Bedoya, Georgetown University, gave the hypothetical example of a recruitment program. "If the program only selects younger applicants, the algorithm will learn to screen out older applicants the next time around" (Brennan, 2015). After California passed laws prohibiting employers from asking about criminal convictions, employers began to use race as a proxy for criminal convictions. As a result, Black applicants received fewer callbacks from employers. The authors concluded that "the real racial gap in felony convictions" is below the level they observed in this field experiment (Agan & Starr, 2018), suggesting that the rules that employers used to determine criminal convictions reflected stereotypes about Blacks as criminals. These types of biases are difficult to uncover because the explanations (e.g., women are nurturing; younger people are adaptable; Blacks are more likely to commit crimes) are based on the very biases they deny (Bonnington, 2018).

We see similar evidence of biases embedded in algorithms when we examine search results. MIT researcher, Latanya Sweeney, conducted a Google AdWords search and found that "black-sounding" names (DeShawn, Darnell, Jermaine) were more likely to lead to search results linking the person to an arrest record than "white sounding" names (Geoffrey, Jill and Emma) (Sweeney, 2013). Datta, Tschantz, and Datta (2015) found evidence that Google used gender to determine the ads to present men versus women; men were shown ads that encouraged them to seek coaching services for higher paying jobs, while women were shown more generic ads. Google Translate interpreted the Turkish phrase "o bir doktor" as "he is a doctor," and "o bir hemşire" as "she is a nurse," equating "male" with the role of doctor and "female" with the role of nurse (Bornstein, 2017). And finally, who is beautiful? According to FaceApp, people with light skin. Light skin continues to be a "benchmark for beauty" among global communities of color. In Asian cultures, white skin is a sign of luxury and prestige; celebrities link their success to lighter skin (Li, Min, Belk, Kimura, & Bahl, 2008). FaceApp reflected these biases in their "hotness" filter, which lightened skin tones to make people more beautiful (Lomas, 2017). What is embedded in these algorithms can reflect or create people's reality.

Further, these algorithms form the backbone of the data brokerage industry. Data brokers ("brokers") compile, aggregate, analyze and sell vast repositories of data about people. Industry estimates place the size of the industry between 2500 and 4000 brokers (Rostow, 2017). Largely unregulated, brokers operate under the radar of the average person, but wield significant power and control over their lives because "even when consumers agree to provide their data for one reason... they rarely have control over how that data will be used later" (Williams et al., 2018). Through the use of proprietary algorithms, brokers cull

data from a variety of sources and compile and aggregate this data for sale. They capture data about virtually every aspect of people's lives: age and race; birth and marriage records; social security numbers; education levels; political views; criminal records; buying habits; health concerns and illnesses; social network connections; etc. (Roderick, 2014). Even Facebook, with all it knows about people, has-purchased data from brokers (Angwin, Larson, Mattu, & Kirchner, 2016).

While the data brokerage industry vigorously defends the privacy of its proprietary algorithms, in an ironic twist, it makes public the intimate details of people's lives. Further, data initially collected for commercial, non-governmental purposes (e.g., targeted marketing) is now being used more broadly (e.g., law enforcement agencies can download digital dossiers from brokers; changes in laws may disclose previously protected information such as preexisting medical conditions or immigration status). It is important that we find ways to make the practices more transparent, "shining the light" on, for example, data brokers' practices (Coates, 2011, pp. 426; Williams et al, 2018).

## LAW ENFORCEMENT & PREDICTIVE POLICING

Predictive policing refers to any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention. Law enforcement agencies have adopted strategies that combine big data and predictive analytics to effectively deploy their resources in areas where crimes are likely to occur. Using historical crime data, police can predict patterns of future crimes - the type of crime, the location, and the likely perpetrator or victim. Predictive policing is assumed to be better than a police officer's hunch because the use of historical patterns of data combined with sophisticated modeling represents a "neutral" approach, one that removes implicit biases that taint human judgement and decision making (Ferguson, 2012). However, as we noted earlier, we must question whether these algorithms remove human biases or more deeply embed these biases in the policing model.

Predictive policing has been used to create profiles of people who possess certain characteristics that are "thought" to be indicative of criminal behavior. For example, young males with arrest records may be more likely to engage in criminal activity than males who are older or males who do not have criminal records. Once we categorize someone (e.g., a criminal), this categorization process affects our perceptions and beliefs about this individual (or group) and tends to "engender bias" (Richeson & Sommers, 2016). Thus, while we can isolate characteristics that help us identify criminals (e.g., arrest records), the characteristics we isolate may be tainted by our biases. In selecting data for inclusion in algorithms, we make assumptions about the characteristics or data to include; these assumptions may also be tainted by our biases.

Algorithms may be tainted in the same manner as our assumptions and may reinforce, rather than override, existing biases. Further, factors such as previous arrest record teach the algorithm to use this data, much like the algorithm in the hypothetical recruitment program learned to select younger candidates (the feedback loop). In this manner, biases become embedded in the algorithm without our awareness or conscious intent. Therefore, when factors such as arrest records, age or other predictive characteristics are used, these factors can reflect human biases rather than neutral statistics, which then calls into question the assumption of neutrality of the mathematical algorithm. Proxies for race (e.g., arrest records) or age (e.g., years of experience) easily reintroduce these factors to the algorithm, even if these factors are explicitly excluded (Buranyi, 2017; Ferguson, 2012; Popp, 2017; Rieland, 2018). When designing algorithms, it becomes important to explicitly identify the presence of potentially biasing factors in order to attempt to minimize their biasing effects (Karppinen et al., 2013; Yetim, 2011).

Researchers at Indiana University (IUPUI) examined bias in predictive policing models, using results from the Los Angeles Police Department (LAPD). They compared differences in arrest rates between predictive policing algorithms and hotspot maps created by LAPD. There was no difference between arrest rates predicted by the algorithm and hotspot maps. However, as the authors note, their analyses did not indicate whether there were biases in arrest rates; the authors could only conclude that the predictive policing models and hotspot maps produced the same results for Blacks and Latinos (Brantingham, Valasik, & Mohler, 2018). But there were questions regarding the objectivity of the study (two of the

authors, Brantingham and Mohler, are on the board of directors of PredPol) and the hotspot maps. Hotspot maps pointed police to specific locations and specific persons ("chronic offender") in certain neighborhoods where police "think" a crime is most likely to occur. People are biased, often through learning or experience, to expect a relationship between certain factors or characteristics (Wason, 1968). So the question here becomes, what have police learned, based on their experience with certain people in certain communities that inform, guide and also bias their judgment and lead to confirmation of their expectation. This in fact was a concern expressed by some community members in the "hotspots;" the factors that might land someone on a hotspot list - contact with the police - might be the result of confirmatory bias (i.e., heavily patrolled spots are more likely to have residents who have higher contact with the police because the areas are heavily patrolled). This "data-driven stereotyping" is entered into the feedback loop that teaches the algorithm about which factors are most predictive of the behavior of interest (Kastle, 2018), thereby subtly reintroducing human biases in the presumably neutral mathematical algorithm.

## DNA PROFILING

The United States has the world's largest DNA database, but they were not the first. In 1995, the United Kingdom created the first national forensic DNA database in the world. Today, there are over 60 countries that store DNA in a database. Members of the European Union provide access to one another's databases (De Moor et al., 2017). In 2009, Kuwait became the first country to institute mandatory DNA testing for all of its citizens/residents (at the time, those who refused could be jailed), but in 2017 mandatory testing was revoked (Coghlan, 2017). DNA databases are widely used.

DNA testing has long been suggested as a useful tool for crime prevention and medical research. DNA databases offer the potential to "collect large amounts of data passively," by mining the databases. The analysis of data gathered passively offers the potential "to improve population health, better understand risk, and potentially reduce health disparities" (Gehlert & Mozersky, 2018). With criminal databases, combing the data "passively" allows researchers or police to find networks of related offenders (e.g., serial offenders) and to identify the "types of criminality in society" (De Moor, Vander Beken, & Van Daele, 2017). Or these algorithms could be used to separate people in ways that are stigmatizing (Drabiak, 2017). Biases may determine the inputs and the populations on which to focus, depending on the questions one seeks to answer.

Law-enforcement agencies have shown a growing interest in mining DNA databases to solve crimes. Indeed, their use in the U.S. helped to solve a recent decades-old case of the "Golden State Killer," a serial rapist and killer who committed crimes from 1974 - 1986. Police matched the unknown killer's DNA sample with samples from the GEDMatch database. While the killer had not uploaded his own DNA sample, one of his relatives had. It was the relative's sample that provided a match to crime scene DNA (Regalado, 2018). But there are nagging questions related to the use of DNA databases. Once someone sends a sample of DNA, the entire family is caught in the web, even if they never provided a DNA sample (O'Brien & Ritter, 2018). Moreover, unfettered access to DNA databases opens the door for studies like Goodwin's which seek genetic causes for behavior.

In the early 1990s, Frederick Goodwin, executive director with the National Institute of Mental Health (NIHM) who led the Violence Initiative Project, posited that the genetic makeup of inner city youth made them more prone to violence, comparing them to "adolescent male monkeys [who] live in gangs and only want to kill each other, have sex, and reproduce." He proposed monitoring infants as young as four months old to determine their propensity for violence and to begin a regiment of drug treatment for those with the potential to be violent (Drabiak, 2017). More recently, Richard Berk, a criminology professor who studies the use of algorithms to predict criminal behavior, is working with researchers in Norway to create an algorithm that will predict, from birth - or before - whether people will commit a crime by their 18th birthday. Critics of his work are concerned that racial biases may be embedded in his models. In one of his algorithms for a U.S. Adult Parole and Probation Department (APPD), he tested a model that excluded race and wealth-related data but included zip codes, which re-

introduced race (and wealth) into his model. While Berk acknowledges that race is a factor, he argues that the models are estimating risk using machine learning rather than people's biases (Popp, 2017), a suggestion that seems to imply that machine bias is less egregious.

There are possible benefits from these analyses, such as early intervention when diseases or problem behaviors are detected early and treated. But there is the question of who defines the criteria that determines the specific behaviors that warrant correction or the diseases to target for treatment - this is where bias can be introduced, explicitly or implicitly. The U.S. response to its recent opioid epidemic illustrates how bias can determine the questions asked and the criteria used to determine decisions or outcomes. Today, the U.S. response to the opioid epidemic has been to label the epidemic as a disease, rather than a crime, and call for treatment options. During a different epidemic, crack/cocaine, the U.S. waged war. Addicts were not viewed as people needing treatment but rather as criminals who should be punished (Cohen, 2015; King, 2017). Algorithms provide the same subtle means to criminalize some, while helping others.

## CONSUMER SCORES & SCORING MODELS

The availability of massive consumer databases has produced a thriving industry of unregulated consumer scores. The World Privacy Forum defines consumer scores as:

"A consumer score that describes an individual or sometimes a group of individuals (like a household), and predicts a consumer's behavior, habit, or predilection. Consumer scores use information about consumer characteristics, past behaviors, and other attributes in statistical models that produce a numeric score, a range of scores, or a yes/no. Consumer scores rate, rank, or segment consumers. Businesses and governments use scores to make decisions about individual consumers and groups of consumers. The consequences can range from innocuous to important. Businesses and others use consumer scores for everything from predicting fraud to predicting the health care costs of an individual to eligibility decisions to almost anything" (Dixon & Gellman, 2014).

While much of the data used to create e-scores used to come from data brokers, increasingly data is coming from mobile and social data. The growth in the number of scores has been exponential. Scores include consumer profitability scores; job security scores; Affordable Care Act health risk scores; medication adherence scores, etc. (Dixon & Gellman, 2014). As Dixon and Gellman (2014) note, the scores in and of themselves are not the problem, rather the concern is with the lack of consumer knowledge or awareness of the scores. This lack of awareness has the potential to wrest control from the individual, placing control in the hands of those who create and control the scoring models. As businesses and governments use these scores to make decisions about people's lives, it is unclear who is examining the biases that may impact the decisions resulting from these scoring models.

In writing about the panopticon, Foucalt (1995), described a laboratory for carrying out experiments to train people to engage in "correct" behaviors. China's Social Credit System points to a modern day example of a panoptic-like algorithm. Similar to FICO credit scores in the U.S., China's scoring model ("Social Credit System") provides an overall measure of a person's worth. The algorithmically determined "sincerity" score defines people's lives. People are evaluated based on five factors: credit history; fulfillment capacity; personal characteristics; behavior; and, interpersonal relationships. Further, they are ranked relative to others; one's rank is publicly available. For example, if you are someone who buys diapers frequently, you are likely to be judged a parent who "is more likely to have a sense of responsibility" versus someone who may be deemed "idle" because they play video games for several hours a day. People with low scores are punished with things like slower internet speeds; restricted access to public venues; restrictions on domestic and foreign travel; limited access to jobs; restricted ability to access social security benefits. High scores, which are equated with good behavior, may result in benefits such as shorter wait times at hospitals; high scores are also a status symbol. To raise one's score, people can inform on their neighbors (Botsman, 2017).

To determine someone's score, China uses "readily observable and easily measurable acts as proxies for behaviors." The vague language of proxies makes it easier to introduce factors assumed to be

excluded from the scoring algorithm. Even more troubling is the possibility that observable proxies raise the specter of social control. One of the proxies for good behavior is support for the scoring system; therefore, one way to raise one's score and ensure access to benefits is to support the scoring system (to do otherwise risks lowering one's score). Further, the model was initially tested in Guizhou province, one of China's poorest provinces (Zeng, 2018), raising the specter of lumping "stigmatized" (poor) people into groups that further marginalize them.

India recently introduced Aadhaar,"the largest biometric identification database in existence." The database uses "photographs, fingerprints, and iris scans to create an ID, linked to social services, including welfare benefits and pensions." Food rations are allocated to the poor via this system. The idea of automating distribution of welfare benefits or food rations suggests the potential for ensuring benefits reach those most in need. Unfortunately, the opposite has occurred. According to Swati Narayan, a researcher for Right to Food, there were a "wave of deaths" among welfare recipients after they were enrolled in Aadhaar. Those who died included Muslims, Dalits, and other "stigmatized" populations. The Indian government continues to test the system (among their poorer populations), leading some to suggest that the most vulnerable are "like guinea pigs" (Bhatia, 2018). Virginia Eubanks (2017), author of Automating Inequality, voices a similar concern, when she argues that big data has been used to profile and control the poor, worsening their conditions.

Contrast the experience of these more vulnerable populations with the experience of a group of Wall Street executives who found themselves under surveillance on their jobs. JP Morgan Chase was a pilot for a surveillance product designed for financial services sectors by Palantir (Winston, 2018). JP Morgan Chase executives raved about the software that "turned data landfills into goldmines" until they discovered that the software was being used to spy on them. JP Morgan Chase's internal "special ops" group used an algorithm to uncover changes in patterns of employees' behavior (e.g., late arrival or early departure) that might suggest "abuse of corporate assets," In response, employees planted fake information in their internal communications to see what was tracked. Employees discovered that the tracking and surveillance reached the highest levels of the firm. As a result, the senior level executives stopped the use of the software (Waldman, Chapman, & Robertson, 2018). Unlike vulnerable populations (e.g., those in China's Guizhou province or India's welfare recipients), these senior level executives were able to terminate the internal algorithm used to secretly monitor and score their behavior.

Baruh and Popescu (2017) argued that algorithms have created "new power imbalances" that create a "privacy discrimination where the personal data of the have nots is disparately and more extensively exploited" than the lives of the haves. We see an example of this when we compare the people targeted in India and China, as well as those targeted with the policing models. However, as the JP Morgan Chase example illustrates, no one is immune from the reach of the algorithms.

## CONSUMER MARKETING

As the previous examples have illustrated, there are benefits as well as risks from the use of algorithms to profile and predict people's behavior. These next examples highlight some of the benefits of data mining in the form of personalized and customized products and promotions. Amazon, one of the pioneers in personalized customization, uses people's data to zero in on present and future needs; indeed, so confident was Amazon that they know their consumers, they filed a patent for "anticipatory shipping," which would ship products before people ordered them (they are currently working out the specifics of this service). There is the oft-cited example of Target's (the big box retailer) use of the predictive power of data to determine if a female customer was pregnant as well as the term of her pregnancy. Because the pregnancy score made people feel uncomfortable, Target ultimately disguised their use of this algorithm by embedding pregnancy promotions among other promotions (Duhigg, 2012). While people became aware of Target's efforts because of their targeted promotions, in most cases, people are not aware of the scoring and modeling that occurs. However, as both of these examples illustrate, clearly there are benefits from mining people's data in the form of personalized and customized products and promotions.

Yet as with many of the applications discussed earlier, there is also a downside. Amazon offered a same day delivery service, Amazon Prime, to certain neighborhoods based on income and zip code. Amazon was accused of "prime-lining" because their service excluded low-income minority neighborhoods. "Low income" turned out to be a proxy for race; thus Amazon's initial selection based on neighborhood income and zip code excluded minority neighborhoods. "Prime lining" is not evidence of intentional bias, but rather an illustration of how easy it is to engage in biased treatment of consumers when factors such as race creep into an algorithm, even when initially excluded (Bornstein, 2017; Bruckner, 2018). In 2008, a businessman returned from his honeymoon to find that American Express had lowered his credit limit from $10,800 to $3,800. He had shopped at a store whose customers (like him) were predominately African American and who had (unlike him) a "poor repayment history." American Express' algorithm profiled the businessman and predicted his payment behavior based on the behavior of other shoppers who fit his profile (Bruckner, 2018; Hurley & Adebayo, 2016). These examples illustrate how algorithms can be biased, by design or by accident, and they also illustrate the difficulty in detecting bias. This is the "darker side" of algorithms that Newman (2014) referred to; information like zip code or income may provide a proxy for race or factors not explicitly included in the algorithm but factors that can invisibly embed biases in the algorithm.

## DISCUSSION & CONCLUSION

More data provides stronger signals that allow companies to better "monetize our social interactions, harvest our data, and render us predictable" (Lewis, 2017). This growing repository of data allows those who control the data to explore common patterns that emerge and, based on those patterns, to identify behavioral traits common among certain groups of people but not others. Once identified, these traits may be used to predict people's future behavior. However an important question to consider is, do the models predict in a way that perpetuates existing biases? It is important to acknowledge the subtle (and not so subtle) ways in which biases are embedded in algorithmic models. When we fail to acknowledge the role of biases, we foster a climate where people may be marginalized in ways that are difficult to detect. Further, if we fail to fully understand the persistence of bias, we fail to fully appreciate the real consequences for real people.

We urge caution, not in the use of predictive algorithms, but in the interpretation of the outcomes and the decisions that result from these models. As these models become more sophisticated, they learn; thus, we must ensure that as with our students, we assess what it is the models have learned and that we test their predictive accuracy. To increase the likelihood of creating more bias-sensitive models, we should seek to bring more diverse voices to the design of algorithms. Moreover, we should seek to more closely examine algorithmic decisions to uncover potential harms that may be heaped on vulnerable populations due to hidden biases. As Friedman, Fireworker, and Nagel (2017) note in their discussion of biases in research, "The bottom line is ... how easy it is for bias - conscious or unconscious - to distort the findings."

As these examples illustrate, the "richly detailed datasets" that exist today provide several benefits, such as more accurate prediction, early warning signs of emerging trends, and customized and personalized product and service offerings. While the outcomes of these models produce clear benefits, these same models can inflict harm on unwitting people. The opaque, black-box nature of the algorithms may disguise biases that drive decisions that on their surface appear neutral because they result from unbiased numbers. Yet, there is evidence to suggest that, whether intentional or not, the biases that exist in these models are hidden under a cloak of invisibility. What gets included in the algorithmic models may be what is most readily available and is often difficult to detect, yet their inclusion may result in decisions that marginalize people (Aguirre, 2000; Bornstein, 2017; Coates, 2011).

Tristan Harris (2016), a former Google design ethicist, in sharing his experience as a magician said this, "Magicians start by looking for blind spots, edges, vulnerabilities and limits of people's perception, so they can influence what people do without them even realizing it." Algorithms perform a similar type of magic.

## FUTURE DIRECTIONS

As the examples in this paper illustrate, biases may be embedded in algorithmic models. "Discrimination is an increasing concern [when we use algorithms] and it really does not matter if the algorithm intentionally or unintentionally engages in discrimination; the outcome on the people who are affected is the same" (Datta, Tschantz, & Datta, 2015). Thus we need to focus on outcomes rather than the intent to engage in discrimination. Datta et al. (2015) suggest that tools be designed to identify and address the biased outcome.

While biases affect everyone, the impact is not the same; biases place a heavier burden on members of marginalized communities (e.g.,racial and ethnic groups, those who are poor). Because these biases operate covertly, rather than overtly, their impact may be harder to see. To fully uncover their impact, we may begin by exploring the "consequences and people" rather than the "processes and technical details" (Gangadharan & Niklas, 2018). Asking people about their experiences not only provides people with a channel to voice their experiences, but also sheds light on some of the real consequences of algorithmic bias. Their self-reports provide insights that may not be captured from an examination of the big data itself. However, self-reports may fail to capture the invisible influences of algorithms. Indeed, that is one of the key challenges explored in this paper, making visible that which is difficult to see.

Combining self-report methods with the "processes and technical details" underlying an algorithm may allow researchers to explore insights gleaned from directly engaging people about their experiences. Welles (2014) argues that there is a tendency in scientific research to use majority populations as the default norm against which minority populations or outliers are compared. As Welles notes, this leads to a tendency to describe anyone who does not fit the norm as deviant. Welles argues that we should examine minority experiences as "reference categories" of their own. This approach may lead to more accurate models and theories. Her arguments support the idea of starting with the voices of those impacted by algorithmic decisions. While everyone is impacted by algorithms, a 2013 U.S. Senate Committee analysis found that the misuse of people's data was more pronounced among the "poor, infirm, elderly, or members of a racial or ethnic minority group." Thus, capturing their voices may provide researchers with insights about new avenues to explore when examining the "processes and technical details" that underly an algorithmic model.

## REFERENCES


A Review of the Data Broker Industry: Collection, Use and Sale of Consumer Data for Marketing Purposes. (2013, December 18). *Staff Report for Chairman Rockefeller. Office of Oversight and Investigations, Majority Staff.* United States Senate, Committee on Commerce, Science and Transportation.

Agan, A., & Starr, S. (2018). Ban the box, criminal records, and racial discrimination: A field experiment. *The Quarterly Journal of Economics*, 133(1), 191–235. doi:10.1093/qje/qjx028.

Aguirre, A., Jr. (2000). Academic storytelling: A Critical Race Theory story of affirmative action. *Sociological Perspectives*, 43(2), 319 - 339.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict criminals. And it's biased against blacks. *ProPublica*. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Baruh, L., & Popescu, M. (2017). Big data analytics and the limits of privacy self-management. *New Media & Society*, 19(4), 579- 596. https://doi.org/10.1177/1461444815614001

Bhatia, R. (2018). How India's welfare revolution is starving citizens. *The New Yorker*. Retrieved from https://www.newyorker.com/news/dispatch/how-indias-welfare-revolution-is-starving-citizens?mbid=social_twitter.

Bonnington, C. (2018). Why it matters that Alexa and Google Assistant finally have male voices. *Slate*. Retrieved from https://slate.com/technology/2018/05/alexa-google-finally-are-getting-male-virtual-assistants-its-about-time.html.

Bornstein, A. M. (2017). Are algorithms building the new infrastructure of racism? How we use big data can reinforce our worst biases—or help fix them. *Nautilus*. http://nautil.us/issue/55/trust/are-algorithms-building-the-new-infrastructure-of-racism.

Botsman, R. (2017). Big data meets big brother as China moves to rate its citizens. *Wired*. Retrieved from http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion.

Brantingham, P. J., Valasik, M,. & Mohler, G. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and Public Policy*, 5(1), 1-6, DOI: 10.1080/2330443X.2018.1438940.

Brennan, M. (2015). Can computers be racist? Big data, inequality, and discrimination. *Ford Foundation*. Retrieved from https://www.fordfoundation.org/ideas/equals-change-blog/posts/can-computers-be-racist-big-data-inequality-and-discrimination/

Bruckner, M.A. (2018). The promise and perils of algorithmic lenders' use of big data. *Chicago-Kent Law Review*, 93(2), 3 - 60.

Buranyi, S. (2017). Rise of the racist robots – how AI is learning all our worst impulses. *The Guardian*. Retrieved from https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses.

Coates, R. D. (2011). *Covert Racism: Theories, Institutions, and Experiences*. Leiden: Brill.

Coghlan, A. (2017). Kuwait's plans for mandatory DNA database have been cancelled. *New Scientist*. Retrieved from https://www.newscientist.com/article/2149830-kuwaits-plans-for-mandatory-dna-database-have-been-cancelled/.

Cohen, A. (2015). How white users made heroin a public-health problem. *The Atlantic*. Retrieved from https://www.theatlantic.com/politics/archive/2015/08/crack-heroin-and-race/401015/.

Datta, A., Tschantz, M.C., & Datta, A. (2015). Automated experiments on ad privacy. A tale of opacity, choice, and discrimination. *Proceedings on Privacy Enhancing Technologies*, 2015(1), 92 -112.

De Moor, S., Vander Beken, T., & Van Daele, S. (2017). DNA databases as alternative data sources for criminological research. *European Journal On Criminal Policy & Research*, 23(2), 175-192. doi:10.1007/s10610-016-9327-9.

Dixon, P., & Gellman, R. (2014). The scoring of America: Secret consumer scores threaten your privacy and your future. *World Privacy Forum*. Retrieved from http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf

Drabiak, K. (2017). Caveat emptor: How the intersection of big data and consumer genomics exponentially increases informational privacy risks. Health Matrix: *Journal Of Law-Medicine*, 27(1), 143-183.

Duhigg, C. (2012). How companies learn your secrets. *The New York Times*. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0.

Eubanks, V. (2017). *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Ferguson, A. G. (2012). Predictive policing and reasonable suspicion. *The Emory Law Journal*, 62, 259 - 325.

Foucalt, M. (1995). *Panopticism. In Discipline & Punish: The Birth of the Prison*. Retrieved from https://foucault.info/documents/foucault.disciplineAndPunish.panOpticism/.

Friedman, H. H., Fireworker, R. B., & Nagel, H. (2017). Biasing research: Can science be trusted? *Journal of Leadership, Accountability and Ethics*, 14(2), 105 - 116.

Gandy, O.H. (2016). Surveillance and the formation of public policy. *Surveillance & Society*, 15(1), 156 - 171.

Gangadharan, S. P., & Niklas, J. (2018). Between antidiscrimination and data: Understanding human rights discourse on automated discrimination in Europe. *London School of Economics*. Retrieved from https://www.researchgate.net/publication/325348206_Between_Antidiscrimination_and_Data_Understanding_Human_Rights_Discourse_on_Automated_Discrimination_in_Europe

Gehlert, S., & Mozersky, J. (2018). Seeing beyond the margins: Challenges to informed inclusion of vulnerable populations in research. *The Journal of Law, Medicine & Ethics*, 46(1), 30-43 https://doi-org.proxy.uchicago.edu/10.1177/1073110518766006.

Greenwald, A. (2017). An AI stereotype catcher. *Science*, 356(6334), 133 - 134. http://science.sciencemag.org/content/356/6334/133

Harris, T. (2016). *How technology hijacks people's minds - from a magician and Google's design ethicist*. Retrieved from (http://www.tristanharris.com/essays/

Helbing, D., et al. (2017). Will democracy survive big data and artificial intelligence? *Scientific American*. Retrieved from https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/.

Hurley, M.& Adebayo, J. (2016). Credit scoring in the era of big data. *Yale Journal of Law & Technology*, 18(2), 148 - 216.

Karppinen, P., & Oinas-Kukkonen, H. (2013). Three approaches to ethical considerations in the design of behavior change support systems. S. Berkovsky and J. Freyne (Eds.): *Persuasive* 2013, LNCS 7822, pp. 87–98.

Kastle, M. (2018). How data analysis is driving policing. *All Things Considered*. Retrieved from https://www.npr.org/2018/06/25/622715984/how-data-analysis-is-driving-policing.

King, N. (2017). Why is the opioid epidemic overwhelmingly White? *All Things Considered*. Retrieved from https://www.npr.org/2017/11/04/562137082/why-is-the-opioid-epidemic-overwhelmingly-white.

Lewis, R. (2017). *Under Surveillance: Being Watched in Modern America*. Austin, TX: University of Texas Press.

Li, E. P.H., Min, H. J., Belk, R. W., Kimura, J., & Bahl, S. (2008). Skin lightening and beauty in four Asian cultures. *Advances in Consumer Research,* Vol. 35. Eds. Angela Y. Lee and Dilip Soman, Duluth, MN:Association for Consumer Research, pp. 444 - 449. http://www.acrwebsite.org/volumes/13415/volumes/v35/NA-35.

Lomas, N. (2017). FaceApp apologizes for building a racist app. *TechCrunch*. Retrieved from https://techcrunch.com/2017/04/25/faceapp-apologises-for-building-a-racist-ai/.

Newman, N. (2104). How big data enables economic harm to consumers, especially to low-income and other vulnerable sectors of the population. *Journal of Internet Law*, 11-23.

O'Brien, M., & Ritter, M. (2018). AP explains: A look at DNA-sharing services and privacy. *Chicago Tribune*. Retrieved from http://www.chicagotribune.com/bluesky/technology/sns-bc-us--ap-explains-dna-databases-20180427-story.html.

O'Neil, C. (2016). *Weapons of Mass Destruction: How Big Data Increases Inequality and Threatens Democracy*. U.S.: Penguin Books.

Popp, T. (2017). Black box justice. *The Pennsylvania Gazette*, 38-47.

Rayome, A. D. (2018). 5 eye-opening statistics about minorities in tech. *TechRepublic*. Retrieved from https://www.techrepublic.com/article/5-eye-opening-statistics-about-minorities-in-tech/.

Regalado, A. (2018). Investigators searched a million people's DNA to find Golden State serial killer. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/611038/investigators-searched-a-million-peoples-dna-to-find-golden-state-serial-killer/

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, 67, 439 -463. doi:10.1146/annurev-psych-010213-115115

Rieland, R. (2018). Artificial intelligence is now used to predict crime. But is it biased? *Smithsonian*. Retrieved from https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/.

Roderick, L. (2014). Discipline and power in the digital age: The case of the U.S. consumer data brokerage industry. *Critical Sociology*, 40(5), 729 - 746.

Rostow, T. (2017). What happens when an acquaintance buys your data?: A new privacy harm in the age of data brokers. *Yale Journal on Regulation,* 34, 667 - 707.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355–374. doi:10.2307/1882010

Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, 56(5), 44-54. doi 10.1145/2447976.2447990.

Varshney, L. R., & Varshney, K.R. (2017). Decision Making With Quantized Priors Leads to Discrimination. *Proceedings of the IEEE*, 105(2), 241–255. doi:10.1109/ JPROC.2016.2608741.

Waldman, P., Chapman, L., & Robertson, J. (2018). Palantir knows everything about you. *Bloomberg*. Retrieved from https://www.bloomberg.com/features/2018-palantir-peter-thiel/.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3), 273 - 281. https://doi.org/10.1080/14640746808400161

Welles, B. F. (2014). On minorities and outliers: The case for making big data small. *Big Data & Society*, 1(1). doi:10.1177/2053951714540613.

Williams, B. A., Brooks, C. R., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions and policy implications. *Journal of Information Policy*, 8, 78 - 115. https://www.jstor.org/stable/10.5325/jinfopoli.8.2018.0078

Winston, A. (2018). Palantir has secretly been using New Orleans to test its predictive policing technology. *The Verge*. Retrieved from https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd.

Yetim, F. (2011). A set of critical heuristics for value sensitive designers and users of persuasive systems. *European Conference on Information Systems (ECIS)* 2011 Proceedings, 185.

Zeng, M. J. (2018). China's social credit system puts its people under pressure to be model citizens. *The Conversation*. Retrieved from https://theconversation.com/chinas-social-credit-system-puts-its-people-under-pressure-to-be-model-citizens-89963.