Intersect, Vol 15, No 3 (2022)

Face Value: How Human Influence Plays a Role in Perpetuating Bias Within Human-Algorithm Interactions

Alice B. Zhang
Stanford University

Abstract

Bias is an inevitable part of the human experience, and it shapes how we think and behave. Unfortunately, bias is not beneficial in many situations. When humans interact with technology, they tend to translate this bias across each exchange. Algorithms, which do not have the privilege of understanding the consequences of bias, may then exacerbate the issue and establish a positive feedback loop that continues to discriminate against certain groups of people. With the increase in human-computer interactions, then, comes an increased need to address this issue. In this study, using a qualitative methodology incorporating the synthesis of past research, I analyze the role of human influence as a designer and as a user and discuss the impact each has on the feedback loop with a particular focus on the recidivism score algorithm COMPAS. I discovered that both explicit designer choice and intrinsic underrepresentation of certain groups in past research play a part in designer influence, while users can influence bias through how they interpret algorithmic results (in which they are either unaware of the issue or value convenience over objectivity). The complex nature of this situation, however, must still be taken into account; there is no single, conclusive reason for human behavior in these interactions, and this study should not determine one perspective as more valid than the other. In that manner, a reasonable way to begin addressing this issue is by first addressing its overarching theme—the lack of careful thinking surrounding algorithmic influence and interpretation.

Introduction

With the heavy integration of technology in modern life comes a greater dependence of humans on machines–and machines on humans. This growing presence of human-computer interaction, then (and the nuances that come with this relationship), is worth examination and comes with its own expectations.

Algorithms, for one, are expected to be perfect—perfect enough to manage the imperfections of humans (Madhavan & Wiegmann, 2007). In other words, they are to be objective where humans are not. This idea,

however, must be questioned on the notion that biased humans can create purely objective algorithms, a concept quite widely believed given the automated quality of technology in the face of human irrationality. After all, algorithms are purely manufactured: their categorization of values lies in a purposely programmed grouping of lists and numbers. Such entities, then, do not hold the genuine ability to value, and entities that cannot value do not have a sense of morality (Véliz, 2021). This idea renders algorithms inanimate and introduces a cruciality to sentiency–a form of consciousness humans possess, but algorithms do not. Let me adopt DeGrazia's definition of sentiency as the ability to have subjective experiences. This incorporation of subjectivity leads to different weightings of interests, and these varying interests consist of the basics of moral status (2020). If algorithms do not possess sentience, their interactions must reflect human sentience.

The subjectivity regarding sentiency, then, can be labeled as another term: implicit perception (and its corresponding biases). Let us define implicit perception as the thoughts and impressions that influence an individual's behavior without their conscious awareness of these influences (Reingold & Ray, 2006). The impulsiveness of these perceptions leads to the formulation of heuristics, which shorten decision-making time and allow for daily functionality without the need to deliberate each thought. The lack of deliberation that accompanies quick thinking, however, leaves the process of heuristics unconsciously incorporating biases into human perception (Cherry & Gans, 2022). Algorithms reflect human sentience. If human sentience is considered implicit perception, and implicit perception leads to bias, algorithms must reflect human bias. In that case, human bias can affect an algorithm's design and training enough that the algorithm itself can be classified as biased.

These algorithms can then influence humans. Laypeople tend to follow advice or results they believe came from an algorithm than another person (Logg et al., 2019), perpetuating more systemic bias and creating a positive feedback loop. An interesting point of investigation, then, is the concept of bias and how it is involved in human-algorithm interactions.

With that in mind, it is more important to determine how human contributions feed into bias than algorithms–human sentiency is more responsible for its perpetuation, and human consciousness allows for a resulting awareness in mitigating the issue. In that case, I want to explore the extent of responsibility humans hold in the instigative and reactive positions of the human-algorithmic bias loop.

The first viewpoint of my research claims that humans are responsible for acting as both the instigator and the reactionary—human bias on the part of an algorithm designer influences algorithmic bias, and human bias on the part of the user then interprets algorithmic results in a biased manner. The second viewpoint reasons that humans may influence algorithmic bias but should be considered victims to its output. With the

two main points of my research identified, I raise my research question: in what ways do humans contribute to the perpetuation and consequences of bias in modern human-algorithm interactions?

Contextualizing COMPAS

Much of my discussion will revolve around the controversy regarding COMPAS, a recidivism score predictor that estimates how likely someone is to commit a crime after release. The higher the recidivism score, the more likely it predicts someone will commit a crime.

In that regard, critics have pointed out COMPAS' habit of assigning high recidivism scores to people of color and lower scores to White defendants. In their May 2016 article, ProPublica compared the petty crime cases of 18-year-old Brisha Borden, who is Black, and 41-year-old Vernon Prater, who is White. Borden and her friend were late to pick up her god-sister and grabbed a nearby unlocked bicycle and scooter to travel more quickly, dropping them when a woman came running after. The previous summer, Prater was charged with shoplifting almost 90 dollars from a Home Depot. Despite Prater's more seasoned criminal history (which involved armed robbery convictions), COMPAS rated Borden as high risk (8 out of 10) and Prater as low (3 out of 10). This prediction was eventually proven false with Borden's clean record and Prater's new eight-year prison term from a more extensive robbery (Angwin et al., 2016). Other cases involve Dylan Fugett/Bernard Parker and Gregory Lugo/Mallory Williams, in which Fugett and Lugo both had subsequent misdemeanors while Parker and Williams did not. COMPAS, however, labeled Fugett and Lugo with scores under three inclusive and Parker and William with scores above six inclusive, with a racial difference between the two pairs aligning with Borden and Prater (Angwin et al., 2016). It has also been found to overclassify women into higher risk groupings than men (Hamilton, 2019).

Unsurprisingly, COMPAS boasts around a 20 percent accuracy rate for predicting who would commit violent crimes after release. When examining its history with a full range of crimes, we recognize an accuracy rate of 61 percent for individuals who re-offended within two years–slightly more accurate than a coin flip (Angwin et al., 2016). COMPAS is, in fact, no more accurate or fair than predictions by people with little to no criminal justice expertise who have their own predispositions against certain groups of people (Dressel & Farid, 2018).

Despite such statistics, many states have encouraged the usage of risk assessments, including COMPAS, in fundamental roles in the courtroom, including states like Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington, and Wisconsin (Angwin et al., 2016). Its increased usage in key roles regarding humans, then, justifies an investigation over the relationship COMPAS has with the perpetuation of bias in the hopes of mitigating future false accusations.

Humans as the Instigator

Algorithms are, by definition, impersonal. It is this impersonality, however, that makes them susceptible to gross data and generalized assumptions (Rainie & Anderson, 2017). The reflexive nature of algorithms renders them as nothing but tools; the agents responsible for these tools, then, are humans (Véliz, 2021). The redirection of responsibility consequently points at a human contribution as their designer towards this susceptibility. It then becomes critical to analyze the role humans play in the perpetuation of bias as the instigator.

Data Selection. Objective algorithms, in the technical sense, are not nondiscriminatory in the way that they can differentiate between unbiased and biased input; rather, they are impartial to all data. In that manner, algorithms are subject to the "garbage in, garbage out" limitation, in which any algorithmic decision can only be as good as the data humans train it on ("BIG DATA," 2016).

In simple terms, objective data leads to objective technology, and biased algorithms reflect biased data. If neutrality is the ideal, it is critical to choose unbiased datasets that will lead to optimal results and circumstances. Unfortunately, much of past data finds itself incorporated within some biased context; non-purposeful data selection, then, can lead to unwanted discrimination against certain groups of people. We can detect this issue in several situations, such as with the predictive policing algorithm PredPol. Police and crime reporters have historically targeted people of color and lower-income classes, so when this information was input into PredPol's training, it perpetuated this discrimination and encouraged police to scrutinize those groups even more (Sankin et al., 2021). In a similar case, both Amazon and Autodesk found their hiring algorithms to be discriminating against women candidates due to being trained on historical data that had a preference for male software engineers (Köchling & Wehner, 2020). One Autodesk recruiter who engaged with more male applicants than female applicants during the first few rounds of hiring even reported how their recruiting AI had learned to give him results similar to the profiles he initially visited, which were unsurprisingly mostly male (Byrne, 2018). With consequences that purposefully harm select groups of people, these examples highlight a certain negligence of algorithm designers in their scrutiny of analyzing input data.

The data fed to COMPAS to train it on recidivism profiling is also proof of an algorithmic reflection towards bias through negligent data selection. COMPAS, as mentioned above, is a recidivism predictor that determines how likely someone will be to commit a crime after release. Researchers found that it labeled a disproportionate number of Black defendants as high risk (Angwin et al., 2016). This result recognizes that COMPAS draws upon past recidivism data, which have roots in past criminal and policing statistics. Given the historical human bias integrated into these datasets, in which police consistently stopped, handcuffed,

searched, and arrested considerably more Black citizens than White citizens (Hetey & Eberhardt, 2018), previous recidivism data would build upon such information and reflect this trend. Simply put, the data COMPAS functioned on mirrored this history of bias against people of color. The algorithm then incorporated these biases into its developed responses, outputting bias against marginalized communities that contribute to the feedback loop of bias.

Direct Design. The above discussion, albeit important, assumes a prioritization of responsibility for dataset selection over other possible causations. This conclusion disagrees with another approach, which claims that the design of the algorithm itself can actively contribute to harm rather than purely acting as a passive reflector of data bias. On that wave, algorithms are not impartial to all data, and some design choices are better than others (Hooker, 2021). In other words, direct human bias on the part of the designer can influence an algorithm to interpret data in a non-objective manner through their concrete design choices. This possibility, then, justifies a closer examination of human interference in algorithm creation. Considering the likelihood of finding unbiased data out of a predominantly biased context, even, we recognize this prospect to be less improbable as an avenue for a solution.

Designers in this perspective are more involved with an algorithm's resulting susceptibility as they directly contribute to the interaction between the given data and the algorithm's response. One example of this involvement could be a creator's arbitrary weighting of certain variables an algorithm would consider to reach a decision given context clues., which COMPAS demonstrates in a biased manner. According to ProPublica, some miscalculations of its recidivism scores stemmed from an inaccurate weighing scale. For example, someone who molested a child could be categorized as low risk because they have a stable job, while someone caught for public intoxication could be labeled high risk because they lack a secure living space (Angwin et al., 2016). In that manner, it weighs stability factors on a higher level than the severity of the crime, which at times can be less relevant to recidivism and therefore erroneously influence the score.

No matter the human effect on an algorithm, why do designers seem to ignore the mitigation of these consequences when creating these technologies? A study examining how people rate algorithm fairness discovered that users tend to rank an algorithm as more fair when it outputs a result in their favor, so much so that any effects of these algorithms on other demographic groups are unseen or ignored (Wang et al., 2020). Given that designers and coders as a whole, even from just a few years ago, were mainly White males, it is likely that any bias in an algorithm was overlooked due to differences over the idea of fairness for represented and underrepresented groups (Rainie & Anderson, 2017). Designers ergo assumed that an algorithm was objective purely from their

point of view, missing potential misrepresentation and correspondingly introducing algorithmic bias to the system.

Underrepresentation. The previous idea, however, introduces a new perspective that does not place as much blame on designers as other positions; instead, it examines how the confounding factor of general underrepresentation of certain groups in data affects algorithm output to unconsciously but inevitably align with biased stances.

"Technology as an abstract concept functions as a White mythology" (Dinerstein, 2006). In other words, the predominancy of White ideals in the academic sphere renders most scholarly records to reflect the White perspective, which naturally translates into modern technology. This overrepresentation stems from its history in the development of human psychology, where all of its earlier works built on the psyche of human participants exclusively involved in higher academics or living near universities (Henrich, 2020). These works originated from western spheres, and early society only saw White individuals able to participate in such academic levels. Therefore, White beliefs established the foundations of psychology and ignored other cultures. This lack of representation in psychological research, however, was overlooked, and the White, Educated, Industrialized, Rich, and Democratic (WEIRD) mindset was considered the "normal" and "universal" mindset of all humans. This assumption later proved false when researchers failed to replicate early experiments with other countries, eventually highlighting the bias established since the outset of these works (Henrich, 2020). In regards to this situation, the subjective viewpoint of WEIRD individuals is intrinsic to the overall representation of human psychology and its corresponding reactions to certain stimuli.

Research and datasets feed into modern technology. Since artificial intelligence prioritizes human interaction, they take information from psychological work. The overwhelming presence of White beliefs in these works, then, leads to an overwhelming amount of White influence on AI and its algorithms. This racialization of technology towards White perception correspondingly continues to misconstrue underrepresented groups and perpetuate bias in the favor of forming a White utopia (Cave & Dihal, 2020).

The responsibility of designers in contributing to bias, then, is less pronounced in this viewpoint. If underrepresentation is an issue to the core of academic research, the concept of biased algorithms is more of an inevitable consequence than a designer's choice. For risk assessment tools such as COMPAS, the overestimation or underestimation of risk towards underrepresented groups is highly plausible given the longtime omission of risk factors relevant to minority populations and the concurrent inclusion of concepts more specific to White offenders (Hamilton, 2018). In that case, the error rates of predictions in risk assessments–and outputs in algorithms in a general sense–will inescapably be larger than the rates

of WEIRD populations, resulting in rampant misrepresentation and corresponding bias out of a designer's control.

Humans as the Reactionary

Nevertheless, humans play an integral part in algorithm design and, by proxy, the potential perpetuation of bias that comes with these design choices. This assumption, in turn, places scrutiny on the designer's role in this perpetuated feedback loop of bias. This position is not, however, the only position that humans can take in regards to algorithm interaction–the role of humans as a reactionary to algorithm output should also be a potential point of investigation.

Unawareness and Overreliance. To adequately discuss the contributions of human users to bias, we must first distinguish the role of a designer from a user. Proper algorithm creation requires the designer to understand the details of their products, including their flaws. The typical human user, however, is not given this information when interacting with technology. Therefore, they are more likely to categorize the benefits of algorithms as absolute. The very idea of algorithmic decision-making, to the common eye, appeals to being objective because of its basis in mathematics and technology (Woods, 2016), which society commonly touts as intelligent fields. These fields are then correspondingly regarded as objective in the face of decisions. The fact that humans create these technologies, then, goes largely ignored, as many believe that algorithms provide "a better standard against which to compare human cognition itself" (Christian & Griffiths, 2017), which aligns with the transcendence of human creation from human ability.

To maintain this belief, many companies themselves will participate in this narrative of algorithm objectivity to promote their technology. Northpointe, the creator of COMPAS, describes their algorithm as "nationally validated" and "Designed to support objective decision-making," in contrast to the controversy surrounding its accuracy and impartiality. PredPol's website touts their algorithm as able to "Proactively patrol to help reduce crime rates and victimization" despite research finding that it based its predictions off the previous, incorrect biases of police.

In lieu of this continuously established perspective, then, people have a strong tendency to trust and provide a reason for what an algorithm outputs (Shafto et al., 2012). If the result of an algorithm is biased, many people will justify it to themselves and inadvertently contribute to the feedback loop. In turn, the role of human reactionaries in the perpetuation of bias is more a lack of awareness and reliance on algorithms spurred by the objective agenda surrounding these technologies than any other conclusion.

Human interpretations of COMPAS results can therefore be influential factors in court proceedings. Let us take Judge James Babler, who oversaw a case over Paul Zilly for stealing a lawnmower and a few

tools. The prosecutor and lawyer both agreed to a plea deal of a year in county jail and follow-up supervision. Babler, however, thought differently. After receiving COMPAS' score on Zilly, which rated him as high risk for future violent crime and as medium risk for general recidivism, Babler overturned the plea deal and sentenced Zilly to two years in state prison and three years of supervision despite the agreement between the prosecution and defense. Sometime later, however, Babler reduced Zilly's sentence from two years to 18 months, stating that "Had [he] not had the COMPAS, ... [he believes that he] would have given one year [or] six months" (Angwin et al., 2016). We classify this decision as an overreliance on technology, where Babler was easily swayed by the algorithm's output and did not consider more possibilities of causation despite the differing opinion of the prosecutor and defendant over his decision, which should have hinted at the idea of accounting for other factors. As noticed by his actions following his initial judgment, he had deliberated over his choice and recognized the immediacy of his trust in COMPAS' prediction, subsequently correcting it. Had he not eventually thought about the case, however, this decision would have continued perpetuating bias against underprivileged groups on the basis of one judge's overdependence on technology, inadvertently continuing the feedback loop.

As the above case highlights the issue with overreliance, it is now fitting to apply this overreliance towards a more explicit demonstration of the perpetuation of bias that comes with this sort of trust and unawareness. Let us refer back to the Autodesk recruiter and his hiring algorithm. The recruiter had an unintentional tendency to examine caucasian male candidates more than other candidates in the initial rounds of hiring and therefore engaged in biased behavior from the start. When the AI took his history into account and correspondingly recommended more caucasian male profiles, the recruiter thought nothing of it as it aligned with his previous results and expectations. In turn, his unawareness of the problem encouraged his preference for caucasian male applicants, systematically eliminating many other profiles, with females having very low priority until his eventual awareness (Byrne, 2018).

In this instance, the recruiter caused his own bias loop. Nevertheless, we can generalize this situation to other recruiters with hiring algorithms that do not take their behaviors into account but instead analyze past data. As past data aligns with the Autodesk recruiter's behavior (where many job positions sought male employees more than female employees), the same disconnect between male and female applicants applies. In either case, the hiring algorithms have already been fed biased information and will therefore continue to perpetuate this group and gender bias.

Fitting a Narrative. To have a complete sense of obliviousness, however, is quite an extreme for many users. The range of unconscious to conscious bias lies on a spectrum, and it is likely to find more users somewhat cognizant of a biased narrative.

How can these kinds of users contribute to bias? Daniel Kahneman introduces a relationship between awareness and effort, in which users can recognize that they have biases and what they are (2013). This recognition can then potentially translate to mindful action, where some individuals will experience quick thinking but then purposively address and correct their thoughts. Others, however, may be unwilling to take the time and effort to counteract these assumptions. This lack of deliberation, then, leads these users to favor evidence that agrees with their preconceptions and biases or make given evidence fit their expectations (Kahneman, 2013). In other words, they are less careful about mitigating biases, as they prefer the convenient process of quick thinking compared to the work required to reason more objectively. This purposeful justification of biased narratives thus continues the feedback loop in an active manner. Michael Graziano furthers this idea by claiming and identifying the two forms of consciousness humans possess: one having the ability to solve complex problems through careful processing, and the other building simplified and subjective models of others' minds, beliefs, and intentions through rapid assumptions (Graziano, 2015; Graziano et al., 2019). If users who value convenience base their conclusions on short assumptions, they build subjective and incorrect representations of other people and subsequently fail at solving complex ethical issues that instead result in discrimination.

If we refer back to Borden, we can recognize this sort of justification from Judge John Hurley, who is normally more careful when it comes to settling bond money; however, in the case of Borden and her friend, Hurley raised the amount for each girl from the recommended 0 dollars to 1000 dollars each. Hurley claims he has no recollection of the case and cannot remember if the scores influenced his decision (Angwin et al., 2016). His forgetfulness of this situation is a very fitting example of quick thinking, in which individuals who do not reanalyze quick thoughts tend not to have absorbed these perceptions. Given that Borden was a Black woman and the severity of Hurley's reaction, it is reasonable to assume that Hurley's biased assumptions against people of color were in play when he saw their COMPAS scores. Rather than question his own biases or COMPAS' objectivity, he automatically took those outputs as correct and behaved in a manner fitting those biases because it was easier for him to accept and move on instead of considering other explanations, in turn continuing the racial bias held by COMPAS and court decisions.

Police are also nonexempt from preferring this sort of convenience. As established above, predictive policing algorithms such as PredPol base their target areas on previous police investigations, which involve communities of color, immigrants, and other marginalized groups. Overpolicing, however, does not occur solely due to algorithmic outputs--they occur because the police readily respond to these results and voluntarily continue to investigate these communities. One terrifying case of this police readiness occurred in Pasco County, where their predictive policing algorithm kept dispatching police to harass one 15-year-old after

they arrested him once for stealing bikes. Over five months, the police went to his home 21 times and showed up at his gym and his parent's workplace. They also made more than 12,500 similar preemptive interrogations on other unsuspecting people whom their algorithm deemed suspicious. These marginalized families were then further victimized when police began to punish them for minor or unrelated incidents, such as fining a mother when they found chickens in her backyard and arresting a father when they saw his son smoking (McGrory & Bedi, 2020). These incidents occur across many areas and over many groups; marginalized groups that are the subject of overpolicing, however, are the ones who get constantly penalized for these situations. With the number of times the police targeted the same families, it is less likely that police are entirely unaware of the bias in their algorithm and instead prefer to continue regarding those families as the "problem families" rather than addressing the real issue and putting in more effort to conduct more careful investigations.

Conclusion

There is much debate over the contributions of humans as a designer or a user to the positive feedback loop surrounding human-algorithm interactions. After all, is it designer choice or intrinsic underrepresentation that should be held more accountable for influencing algorithmic bias? Is it better to consider users as unaware of their role in bias preservation or as those who value convenience over effort for objectivity?

Despite the specific differences in these analyses, one central theme persists—the lack of careful thinking within each interaction. For designers, we recognize a lack of consideration for what data to input into algorithms or what algorithm designs lead to ethical outputs. For users, we identify a lack of effort in either being aware of the issue or attempting to mitigate it. Many of the situations discussed above arose from this sort of carelessness. Our main goal, then, is to address the importance of purposeful thought for the ethical issues that impact a number of communities.

One step towards this goal could be the enforcement of transparency standards and open-source code for specific AI that ensures that an algorithm is not only understood by the public but also understood by the designer. AI Now, a nonprofit endorsing algorithmic fairness, advocates for the idea that a designer must be able to explain the reasons for an algorithm's decisions before the public can use the algorithm (Byrne, 2018). This idea places pressure on the designer to ensure that they deliberate the way they train or design their algorithms and truly choose what they think is best for their algorithm's ethical performance. New York's city council and mayor have already called for this sort of transparency in response to the controversy surrounding COMPAS uncovered by the ProPublica study (Byrne, 2018).

Designers and companies responsible for the creation of these algorithms can also make an effort to mitigate the issue of underrepresentation by creating diverse teams that can more thoroughly examine and remove the biases found in most of the training data fed to machines. The added diversity can then contribute to the intentional screening of biased correlations in the data, where people of different backgrounds can more easily identify biases regarding their experiences. Data correlating men to the office and women to the kitchen, for example, can be identified with a more female-centric team and removed through careful text-preprocessing of a variety of keywords (Byrne, 2018; Bhavsar, 2018). By placing policies that demand careful decision-making and explicitly encourage representation, designers will be more likely to consider the ethical implications of their behavior instead of neglecting it as an issue for later.

Users, on the other hand, should be purposely reminded of the flaws in technology and required to undergo training that redefines how they interpret algorithmic outputs. Predictive policing systems, for one, could start by acknowledging the errors that can occur when an algorithm makes a decision; by doing so, police departments are more encouraged to personally audit for errors and correct them, which proper training can ensure (Ferguson, 2017). Those in general law enforcement could find success in more conscious thinking by trying to disprove their immediate theories or conclusions instead of proving them (Murgado, 2014). In the context of COMPAS and recidivism, this concept encourages judges to actively reevaluate their immediate conclusions from COMPAS' output before taking action on their decision by ensuring that they mull over the details well enough to be fairly certain whether someone will reoffend or not. Before police follow the result of a predictive policing algorithm, they can engage in this concept to reconsider whether the history of the targeted family warrants another investigation.

No matter the extent of responsibility humans play in the perpetuation of bias, they are still responsible to some degree as both an instigator and a reactionary. That is all that matters. Disregarding the role we play in this feedback loop, we are all humans, and we all have human awareness. Let us take our gained awareness of this issue and the knowledge found in these conclusions to actively mitigate such bias to the best of our ability: it all starts, really, with taking things deeper than face value.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica.
Retrieved January 16, 2022, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bhavsar, P. (2018, October 11). *Making Amazon hiring AI unbiased*. Medium. Retrieved June 13, 2022, from

https://towardsdatascience.com/making-amazon-hiring-ai-unbiased-129c5a2bef14
Byrne, W. (2018, February 28). *Now Is The Time To Act To End Bias In AI*. Fast Company. Retrieved June 12, 2022, from https://www.fastcompany.com/40536485/now-is-the-time-to-act-to-stop-bias-in-ai
Cave, S., & Dihal, K. (2020). The Whiteness of AI. *Philosophy & Technology*, *33*(4), 685–703.
https://doi.org/10.1007/s13347-020-00415-6
Cherry, K. (2022, February 13). *What Are Heuristics?* Verywell Mind. Retrieved March 10,
2022, from https://www.verywellmind.com/what-is-a-heuristic-2795235
Christian, B., & Griffiths, T. (2017). *Algorithms to live by: The Computer Science of Human*
Decisions. Picador.
DeGrazia, D. (2020). Sentience and consciousness as bases for attributing interests and moral
status: Considering the evidence and speculating slightly beyond. *Neuroethics and Nonhuman Animals*, 17–31. https://doi.org/10.1007/978-3-030-31011-0_2
Dinerstein, J. (2006). Technology and its Discontents: On the Verge of the Posthuman. *American*
Quarterly, *58*(3), 569–595. https://doi.org/10.1353/aq.2006.0056
Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism.
Science Advances, *4*(1). https://doi.org/10.1126/sciadv.aao5580
EXECUTIVE OFFICE OF THE PRESIDENT. (2016). (rep.). *BIG DATA: A REPORT ON*
ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS. Retrieved February 19, 2022, from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.
Ferguson, A. G. (2017). Policing Predictive Policing. *Washington University Law Review*, *94*(5). https://openscholarship.wustl.edu/law_lawreview/vol94/iss5/5
Graziano, M. (2015). *Consciousness and the Social Brain*. Oxford University Press.
Graziano, M. S., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2019). Toward a standard model of consciousness: Reconciling the attention schema, Global Workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, *37*(3-4), 155–172. https://doi.org/10.1080/02643294.2019.1670630
Hamilton, M. (2018). The Biased Algorithm: Evidence of Disparate Impact on Hispanics. *56*
AM. CRIM L. REV. 1553 (2019).

Hamilton, M. (2019). The sexist algorithm. *Behavioral Sciences & the Law*, *37*(2), 145–157.
https://doi.org/10.1002/bsl.2406
Henrich, J. P. (2020). *The WEIRDest People in the World: How the West Became Psychologically*
Peculiar and Particularly Prosperous. Farrar, Straus and Giroux.
Hetey, R. C., & Eberhardt, J. L. (2018). The numbers don't speak for themselves: Racial
disparities and the persistence of inequality in the Criminal Justice System. *Current Directions in Psychological Science*, *27*(3), 183–187. https://doi.org/10.1177/0963721418763931
Hooker, S. (2021). Moving beyond "algorithmic bias is a data problem." *Patterns*, *2*(4).
https://doi.org/10.1016/j.patter.2021.100241
Kahneman, D. (2013). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
Köchling, A., & Wehner, M. C. (2020). Discriminated by an Algorithm: A systematic review of
discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*, 795–848. https://doi.org/10.1007/s40685-020-00134-w
Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer
algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.
https://doi.org/10.1016/j.obhdp.2018.12.005
Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human
and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301.
https://doi.org/10.1080/14639220500337708
McGrory, K., & Bedi, N. (2020, September 3). *Targeted*. Tampa Bay Times. Retrieved June 12, 2022, from
https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/intelligence-led-policing/
Murgado, A. (2014, July 17). *Dealing with confirmation bias*. POLICE Magazine. Retrieved June 13, 2022, from
https://www.policemag.com/341175/dealing-with-confirmation-bias
Rainie, L., & Anderson, J. (2017). (rep.). *Code-Dependent: Pros and Cons of the Algorithm Age*.
Pew Research Center. Retrieved January 16, 2022, from
https://www.pewresearch.org/internet/2017/02/08/theme-4-biases-exist-in-algorithmically-organized-systems/.
Reingold, E. M., & Ray, C. A. (2006). Implicit Cognition. *Encyclopedia of Cognitive Science*.
https://doi.org/10.1002/0470018860.s00178

Sankin, A., Mehrotra, D., Mattu, S., Cameron, D., Gilbertson, A., Lempres, D., & Lash, J. (2021,
December 2). *Crime prediction software promised to be bias-free. new data shows it perpetuates it*. Gizmodo. Retrieved February 21, 2022, from https://gizmodo.com/crime-prediction-software-promised-to-be-free-of-biases-1848138977
Véliz, C. (2021). Moral zombies: Why algorithms are not moral agents. *AI & SOCIETY*, *36*,
487–497. https://doi.org/10.1007/s00146-021-01189-x
Wang, R., Harper, F. M., & Zhu, H. (2020). Factors influencing perceived fairness in algorithmic
decision-making. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376813
Woods, T. (2016, June 8). *'Mathwashing,' Facebook and the Zeitgeist of Data Worship*.
Technical.ly. Retrieved February 20, 2022, from https://technical.ly/2016/06/08/fred-benenson-mathwashing-facebook-data-worship/ s