A Perfect Storm for Epistemic Injustice: Algorithmic Targeting and Sorting on Social Media

Heather Stewart
Oklahoma State University
heather.stewart@okstate.edu

Emily Cichocki
Western University
ecichock@uwo.ca

Carolyn McLeod
Western University
cmcleod2@uwo.ca

A Perfect Storm for Epistemic Injustice:
Algorithmic Targeting and Sorting on Social Media
Heather Stewart, Emily Cichocki, and Carolyn McLeod

Abstract

  Over the past decade, feminist philosophers have gone a long way toward identifying and explaining the phenomenon that has come to be known as epistemic injustice. Epistemic injustice is injustice occurring within the domain of knowledge (e.g., knowledge production and transmission), which typically impacts structurally marginalized social groups. In this paper, we argue that, as they currently work, algorithms on social media exacerbate the problem of epistemic injustice and related problems of social distrust. In other words, we argue that algorithms on social media recreate and reify the conditions that lead to some groups being systematically denied the full status of knowers, thereby corrupting the epistemic terrain and, with it, systems of social trust and cooperation. We argue that algorithms do this in two ways—namely, via what we are calling algorithmic targeting and algorithmic sorting.

> "The Web reflects and even amplifies real-world inequities as often as it ameliorates them."
> —Astra Taylor, 2014[1]

> "The problems here are complex, both technically and philosophically."
> —Mark Zuckerberg, 2016[2]

---

[1] Taylor 2014, 10.

[2] This quote from Facebook CEO Mark Zuckerberg was initially a comment Zuckerberg posted on his personal Facebook account in 2016. We borrow the quote from Robert Smith (2019, 16).

1. Introduction[3]

The development, implementation, and future possibilities of AI have recently gained much scholarly and popular attention, and rightly so given the many ethical and social challenges raised by AI (see Coeckelbergh 2020, 3–8). Among these challenges is how AI can cause or worsen social injustice. Most of the current research on this narrower topic has discussed institutional uses of these technologies (e.g., algorithms used in assessing applications for employment, housing, or loans; and algorithmic predictions regarding the likelihood of "recidivism" in criminal justice contexts).[4] Our focus instead is on the impact that algorithms can have at the interpersonal level, especially algorithms deployed in social media. We argue that AI in social media creates conditions that have the potential to worsen *epistemic injustice* (see Fricker 2007; Medina 2013), both on and offline. We further argue that this problem, and a related problem of diminishing social trust, ought to be considered when thinking about the potential risks of rapidly developing AI for oppressed groups in society. In particular, the problems we identify ought to be considered as an important part of growing dialogues surrounding what's been called "algorithmic injustice" and "algorithmic oppression" (see Buolamwini 2016; Eubanks 2018; Noble 2018).

Although our focus is generally on understanding algorithmic injustice or oppression better, and more specifically how algorithmically driven social media platforms can worsen unjust social and epistemic conditions, we acknowledge that social media can (and often does) serve a valuable role in solidaristic efforts at *resisting* social injustice and creating positive social change (Stephen 2015; Tufekci

---

[3] We are grateful to audiences at the CEPE/IACAP Joint Conference on the Philosophy and Ethics of Artificial Intelligence, hosted by the University of Hamburg, as well as the Feminism, Social Justice, and AI Workshop, hosted by the University of Waterloo. Specifically, we are grateful for helpful comments from Judith Simon, Carla Fehr, Karen Frost-Arnold, and Catherine Stinson. Finally, we are grateful for helpful conversations about AI ethics with colleagues at Western University, and specifically those who participated in a reading group on the topic in the 2020–2021 academic year.

[4] We do not deny that the focus on institutional-level uses of algorithms is of critical importance. After all, these uses can have significant impacts on the material conditions of people's lives (e.g., whether they can secure a loan, housing, a job; whether they are let out of jail on parole; etc.). Nevertheless, we contend that there are much broader social and interpersonal implications of our regular engagement with algorithms, which is why we focus on the interpersonal level.

2014, 2017).[5] We do not intend to deny the potential of social media to be used as a force for good.[6] Nevertheless, we believe it is important to attend to the potentially harmful side of widespread, unreflective social media engagement, especially engagement that occurs without awareness or understanding of the underlying algorithmic processes that shape it.

The paper proceeds as follows. In section 2, we cover preliminaries about AI and algorithms, clarifying how we understand them for the purposes of this paper. Specifically, we discuss the role of algorithms on social media sites, such as Twitter, Instagram, and Facebook. In section 3, we describe two different but related algorithmic processes on social media, which we call *algorithmic sorting* and *algorithmic targeting* and which we later tie to epistemic injustice. Section 4 provides an overview of the problem of epistemic injustice, as it has been described by feminist philosophers. We explain the moral and social significance of epistemic injustice, including the issues it poses for trust and effective communication between members of different social groups. In section 5, we argue that algorithmic sorting and targeting on social media create conditions that likely worsen epistemic injustice and accompanying social distrust and degradation of intergroup communication. Finally, in section 6, we discuss the implications of our argument for developing conversations about algorithmic injustice and algorithmic oppression.

---

[5] As Alex Kantrowitz (2020) notes, many social movements (e.g., Occupy Wall Street, the Arab Spring, the Gezi Park protests, and Black Lives Matter) have relied on social media for spreading information and organizing resistance efforts. To give an example current to the time of writing this paper, social media platforms such as Facebook and Instagram are currently being used to keep attention on the Fairy Creek Blockade as Indigenous land defenders protect old-growth forests in unceded Pacheedaht and Ditidaht territory (or the territory known as Vancouver Island, British Columbia). We would like to once again thank the attendants of the Feminism, Social Justice, and AI Workshop for emphasizing this point.

[6] For a discussion of this complexity, see Tufekci's 2014 article, "Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape." There, she examines the complex relationships between social media and activism, noting that most social movements "have thoroughly integrated digital connectivity into their toolkits, especially for organizing, gaining publicity, and effectively communicating," and that while "social media's empowering aspects are real and profound . . . these impacts do not play out in a simple, linear fashion" (Tufekci 2014, 1). Tufekci notes that "social media both empowers new digitally-fueled movements and contributes to their apparent weaknesses in seemingly paradoxically ways" (1).

2. AI, Algorithms, and Social Media

What is meant by "AI" in debates on this topic is not always clear, nor is it consistent. Before moving forward, we want to be clear ourselves about how we are thinking about AI.

Nick Polson and James Scott (2018) argue that when we hear "AI," we shouldn't think of something out of a science fiction fantasy but rather simply of algorithms. Algorithms, they explain, are no more than a set of step-by-step instructions that a computer follows in order to perform some task. "On its own," they note, "an algorithm is no smarter than a power drill; it just does one thing very well, like sorting a list of numbers or searching the web for pictures of cute animals" (Polson and Scott 2018, 3). Put that way, AI doesn't sound nearly as frightening or as complicated as we might intuitively think. Algorithms are simply tools that are oriented toward some particular goal and that function via a simple set of instructions. In order to get algorithms to complete increasingly complicated tasks, algorithms get chained together. The resulting chains of simple algorithms create a "domain-specific illusion of intelligent behavior" and are what we think of as "AI." Almost every AI system follows the same "pipeline-of-algorithms" template. The pipeline takes in a bunch of data from a particular domain, performs a chain of calculations, and outputs some prediction, decision, or information (Polson and Scott 2018, 3–4). Voilà—AI! A tidy definition of AI that we can work with, borrowed from Mark Coeckelbergh (2020, 64), is that AI is "intelligence displayed or simulated by code (algorithms) or machines." Henceforth, this is how we will think of AI.

Many of the digital technologies that we interact with on a daily basis run algorithms. While algorithms do many things, what is important for our present purposes is that they give shape to our informational channels on the internet. They do that, for example, by mediating what appears in our search engines (Noble 2018; Sumpter 2018, 105), by determining what news sources are and are not presented to us (Sumpter 2018, 105), and by structuring what appears on our social media feeds and in what order.

Despite the fact that algorithms play critical roles in our lives, many people do not understand them. Robert Elliot Smith (2019, x) notes that most users are generally "unaware of exactly how [algorithms function], mostly [they] don't understand their operation, and [they] barely grasp the influence they exert on our lives." Smith (2019, 63) goes on to say that our willing but uninformed acceptance of algorithmic operations in our lives results from our assuming that algorithms are benign and unbiased—the result of objective mathematical computation, free of the corrupting influence of social values. It is clear that far more work is needed to educate digital technology users: to help people interact with online platforms with a more critical eye.

As noted above, our focus is on social media, which operates via algorithms that are often invisible to individual users, despite being central to the functioning of these platforms. We are concerned as well with the ethical and social challenges that these platforms pose. David Sumpter outlines just some of these challenges in his 2018 book, *Outnumbered: From Facebook and Google to Fake News and Filter-Bubbles—The Algorithms That Control Our Lives.* About the social media situation around him while he was writing the book, he writes, "Twitterbots were spreading fake news . . . far-right groups were living in algorithmically created filter-bubbles; Facebook was measuring our personalities, and [the data] were being exploited to target voters" (Sumpter 2018, 14). He then continues, "All of the big Internet services—including Google, Yahoo, Facebook, Microsoft, and Apple—build up a personalized picture of our interests and use these to decide what adverts to show us" (19).

The ability of social media to generate a hyperpersonalized picture of a specific user is particularly effective. One programmer notes, "The power of the 'like' button on Facebook to target adverts is scary. Your 'liking' gives a lot of insight into you as a person" (programmer Angela Grammatas, quoted in Sumpter 2018, 19). This information—about who we are as people—is a source of profit for Facebook. That is because the better they know us, the better they can tailor advertisements to us or generate a feed that is more likely to keep us on the site longer (our attention to and engagement with the site is a financial asset to social media corporations like Facebook). It is therefore part of the business model of social media sites like Facebook to get to know us *really* well.[7] The goal is to tailor content to specific users and to keep them engaged and on the platform, at any cost. This shaping of our social media feeds in ways that fit our identities and interests is done, in large part, by algorithms.

3. Algorithmic Sorting and Targeting

The content we see on social media is curated with our "likes" and other background information in mind through what we call "algorithmic sorting" and "algorithmic targeting." We will later argue that these processes are likely to worsen problems of epistemic injustice and social distrust.

---

[7] We will not review how the algorithms utilized by social media sites such as Facebook go about classifying us here, although Sumpter (2018, 26–33) offers a really accessible description of how this works. To understand specifically how algorithms make predictions about our interests based on past "likes," see Sumpter (2018, 105–17).

3.1. Algorithmic Sorting

Algorithmic sorting, as we are using this term, describes the increasing separation of people on social media into different epistemic worlds, which have little overlap in informational content. The result is that people come to know—or think they know—different "facts" and may even rely on different standards of evidence. The worldviews of others become increasingly distant, unfamiliar, or unrecognizable.

Before going into further detail about algorithmic sorting, a point of terminological clarification is in order. Within computer science, a different term—*sorting algorithm*—describes an algorithm that functions to put particular elements of a list in numerical, lexical, or some other specific order. Sorting algorithms are often used to help sort data in ways that help optimize the function of other algorithms, including search algorithms, such as those used on common search engines, like Google. This term—*sorting algorithm*—is a noun. It denotes a particular type of algorithm. What we are describing, on the other hand, is more of a verb: a process that is driven by algorithms of different types and that has the effect of sorting people into qualitatively different epistemic worlds. Bear in mind, moving forward, that our focus is on algorithmic processes that produce certain epistemic and social outcomes, not a particular type of algorithm.[8]

Bearing that terminological clarification in mind, and in order to further explicate the phenomenon of algorithmic sorting, we must now take a detour into literature on two phenomena that are related to algorithmic sorting: echo chambers and filter bubbles.[9] Although most theorists take these two phenomena to be distinct, there is often some conflation of them and slippage in the usage of the terms "echo chamber" and "filter bubble." A common way of distinguishing them is as follows:

> In a paradigm case of an echo chamber, bloggers linked to other bloggers who agreed with them, confirming their views and supporting what they already thought . . . . Each set of bloggers had created their own world, within which their views reverberated . . . . The difference between "filtered" and "echoed" cavities lies in whether they were created by algorithms or by people. While the bloggers chose the links to different blogs, algorithms based on our likes, our web searches, and

---

[8] We are grateful to Catherine Stinson for pushing us to clarify our terminology here, relative to the similar terminology used in computer science.

[9] There has been much discussion of these phenomena in the past decade, following the 2011 publication of Eli Pariser's book, *The Filter Bubble: What the Internet is Hiding from You.* This book generated both discussion and critique; see, for example, Bruns (2019) for an argument that the impact of filter bubbles and echo chambers is overstated.

our browsing history do not involve an active choice on our part. It is these actions which can potentially create a filter bubble. Each action you make in your web browser is used to decide what to show you next. (Sumpter 2018, 137)

On this view, echo chambers are the result of human agency: the user opts into certain informational networks, which were curated by humans and reinforce certain ideas or perspectives. Beliefs "echo" throughout these networks, relatively immune from outside challenge or resistance. This process functions to firm up the beliefs (e.g., confirmation bias goes to work). As a paradigm example, you might think of social media groups, such as Facebook groups or community blogs, in which people self-select into the group, the group contains primarily (or in some cases, exclusively) like-minded others, and the information shared through the group is of the sort that confirms the relevant beliefs and the standards of evidence upon which they are built.

On the other hand, filter bubbles are taken to be the result of algorithms learning about us via a variety of available data inputs (e.g., our "likes," our viewing patterns, and other data traces we leave behind on the web) and curating our informational worlds accordingly (e.g., based on what the algorithm predicts that we most want to see, or what will most likely keep us engaged and online).[10] Notably, on this view, the "filter bubble" isn't a perfectly closed epistemic world, not in the way an echo chamber might be. Not *all* opposing views will be sorted out, at least not right away. However, the filter bubble effect involves the algorithm becoming increasingly good at knowing what information interests you and at tailoring your network to reflect only that information.[11]

While echo chambers and filter bubbles are often treated as though they are distinct, in practice, they are quite difficult to tease apart. In other words, the overlap between human-created echo chambers and algorithmic-driven filter bubbles is quite messy in practice. The dynamic nature of user engagement and algorithmically curated content and choices blurs the line between human agency and algorithmic influence in significant ways.[12] Some of our epistemic world-making online most

---

[10] For a clear and helpful overview of how this works, see Sumpter (2018, 138–51).

[11] This way of parsing out filter bubbles and echo chambers is common. But note that there have been other helpful proposals for understanding these or related concepts, such as "epistemic bubbles." For example, see Thi Nguyen (2020), who demarcates "echo chambers" from "epistemic bubbles" by claiming there is a cultivated, systematic distrust in the case of "echo chambers" that is lacking in the case of "epistemic bubbles."

[12] For a detailed treatment of the relationships between algorithms, autonomy, and agency, see Rubel, Castro, and Pham (2021).

certainly happens via our conscious choices and reflects our (mostly explicit) decision-making (e.g., whose friend requests we accept or deny, which pages we follow, and which groups we opt into joining). In these ways, our social media feeds might come to disproportionately "echo" our extant beliefs back to us and do so as a result of curatorial decisions we have made as users. For example, a queer feminist academic might, as a result of consciously curating her friends list, end up having a network that mostly shares left-leaning news articles and queer feminist content. In this way, as a result of her own actions, many of her extant views end up being reflected (or "echoed") back at her when she views her feed. This is the sense in which human choices are involved in curating our informational spheres.

However, the curation is not merely the result of our agency, since the decisions we make are already framed for us to some extent. For example, our "suggested friends" lists are created and presented to us by algorithms; suggestions of pages to follow and groups to join are curated by algorithms too. Whether we request those suggested friends, accept the request others send us based on their own algorithmic recommendations, or join those suggested pages or groups, is of course our decision, but any sense we have that we are acting *wholly* on our own choices is illusory.[13] The choices are already framed in powerful ways by the algorithms. So, echoing effects aren't the exclusive product of human choice; rather, the human choices are aided by the output of algorithms.

Relatedly, the filtering effects that are seen to be the result of algorithms are informed by active user choices. We are, after all, the ones providing the data to the algorithms, which then filter information based on that data. In other words, it is our decision-making and activity that gives form to the algorithm's predictions and suggestions, or that aids the algorithm in tailoring our feeds more obviously to our interests.

Rather than being entirely distinct, echo chambers and filter bubbles are therefore in a complex feedback loop. On the basis of (already constrained) decisions we make in our online social sphere, the algorithms learn important information about us (e.g., whom we are inclined to add, and what we are inclined to follow, view, or engage with). The algorithms then go to work "filtering" our online social and epistemic sphere accordingly. Information is filtered through the lens of our interests and personalities, echoing back what we have taught the algorithm about us. The way

---

[13] The same is true, of course, of other choices we make, those that are not influenced by algorithms. In particular, these choices are shaped by our social circumstances, which limit the options we have available. It would be interesting to explore whether our options are more (or less) limited than usual on social media and how our autonomy is correspondingly limited. We reserve discussion of this issue for another day.

things appear on our social media feeds is then the result of a messy overlap between algorithms and active choices, or choices moderated by algorithms.

Given this overlap and the difficultly of parsing out what is created by humans vs. algorithms, we opt to use the more general term of *algorithmic sorting*. In other words, rather than refer to echo chambers and filter bubbles, we want to focus on the *processes* that lead us *into* these increasingly separate and polarized epistemic worlds. As we define it, "algorithmic sorting" refers to the sorting of social media users into increasingly different, increasingly closed-off, and deeply biased informational spheres as a result of algorithms and algorithmically constrained human decisions.

3.2. Algorithmic Targeting

The second algorithmic process that we would like to highlight is what we are calling *algorithmic targeting.* Algorithmic targeting on social media occurs when information or content is placed in front of the user based on the algorithm's prediction about their likelihood to view or engage with that content. Algorithms effectively direct users' attention to specific content that is potentially of interest to them based on their previous engagement, "like" patterns, or time spent on particular pages. This process contributes to algorithmic sorting, though the two processes are not coextensive. Algorithmic targeting is unique—and perhaps especially pernicious—because it involves presenting specific content to users, based on their susceptibility to accepting, believing, or approving of it, as determined by the algorithm. In this way, algorithmic targeting plays a special role in epistemic world-making, as it reinforces beliefs that users might already be sympathetic or susceptible to, and which appear "normal" within the epistemic worlds into which they have been sorted. For example, if a person has anti-vax misinformation explicitly targeted to them (e.g., via a recommended video), this can be particularly pernicious once they have been sorted into a world with like-minded others who share in anti-vax beliefs. It can feel as if the content being presented "must be correct" since everyone else in the epistemic world accepts it, too.

Some algorithmic targeting is relatively benign (setting aside anticapitalist concerns we might have about the underlying motivation for targeting things to people in the first place, which is, of course, to generate profit). For example, when Heather opens her Facebook page, the things targeted to her may include advertisements for new camping equipment, event links to a recently announced fall tour for her favorite musical artist, and suggested vegan restaurants in the area. These bits of content are channelled to Heather based on what the algorithm has gleaned about her interests: in camping, in live music, and in plant-based cuisine,

respectively.[14] What this targeting does more than anything, perhaps, is get Heather to act on her preferences: she might buy the camping gear or concert tickets, or get dinner from the new vegan spot. But it can also reinforce beliefs she has about the value or popularity of the things she prefers. The message she may receive, if only subconsciously, is that these things must be important, for why else would advertisements for them be appearing all over her feed? They must, then, be good. To this end, even in the relatively benign cases, algorithmic targeting still plays on our cognitive biases in interesting, and perhaps worrisome, ways.

Some algorithmic targeting is more directly concerned with what we ought to believe (as opposed to what we ought to buy or order for takeout). The recommendations are that we take certain theories or views about the world seriously. For instance, it is now widely documented that social media algorithms target conspiracy theories and radical views at susceptible users. Studies have shown that the algorithms on Instagram push misinformation, dangerous conspiracy theories, and white supremacist content at specific users (Campbell and Holroyd 2021). Facebook suggests radical political groups that propel ideologies such as white supremacy (Paul 2021). On YouTube, the algorithms deliver people "novel and shocking claims," leading people to describe what has been called "the three degrees of Alex Jones," a running joke meant to highlight that no matter where you started on YouTube, you were never more than three recommendations away from a video by the right-wing conspiracist who popularized the idea that the Sandy Hook school shooting in 2012 had never happened and the bereaved parents were mere actors playing parts in a murky conspiracy against gun owners (Tufekci 2018). Amazon, too, is guilty of sending customers toward increasingly extremist misinformation, including white supremacist literature (Long 2021). All these platforms target such content at users by "cross-propagating" from other conspiracy theories; if, for example, someone looks at a lot of anti-vax content and COVID-19 denialism, the algorithms direct them toward increasingly radical conspiracism (Long 2021). On each platform, algorithms are directing people toward hateful and conspiracy-ridden content based on the users' predicted susceptibility (a product of their past engagement, "like" patterns, etc.)—and all for the sake of generating profit for companies (Zeynep Tufekci, quoted in Pierce 2019).

Having now described how we understand algorithmic sorting and targeting, we can consider how they contribute to epistemic injustice. Our first step is to give some preliminaries about epistemic injustice.

---

[14] For a discussion of why targeted advertisements on Facebook are so astoundingly accurate and effectively personalized, see Jennings (2018).

4. Epistemic Injustice

The problem of epistemic injustice has become a focal point for feminist philosophers in recent years, largely following the publication of Miranda Fricker's 2007 monograph, *Epistemic Injustice: Power and the Ethics of Knowing*, and subsequent works that build on the concept (see, e.g., Medina 2013; Kidd, Medina, and Pohlhaus 2017). Although Fricker was not the first to describe the phenomenon now widely known as epistemic injustice (see, for example, Collins 2000; Lorde 1984), she has provided an invaluable framework for understanding it, which we will draw upon here.

As Fricker (2007) explains it, epistemic injustice is an injustice that someone experiences in their capacity as a knower, most typically when their status as a knower has been diminished owing to "negative identity prejudicial stereotypes" that target a social group of which they are a part (i.e., their marginalized race, gender, sexual orientation, socioeconomic status, or the like). Although Fricker identifies two varieties of epistemic injustice—what she calls testimonial injustice and hermeneutical injustice—for our purposes, we will focus on *testimonial injustice.* A subcategory of epistemic injustice, testimonial injustice is an unfairness that someone experiences with respect to the uptake of their testimony. More precisely, they experience an unjust deflation of their credibility owing to stereotypes held about some facet of their identity. For example, an innocent Black male defendant is a victim of testimonial injustice in the courtroom when jurors dismiss his testimony because of the triggering of widely held stereotypes about Black men, such as stereotypes that associate Black men with criminality, aggression, or violence. This is an instance of testimonial injustice because the Black defendant's ability to have his testimony fairly evaluated is blocked due to social stereotypes that the jurors hold (implicitly or explicitly) and that are widespread in the broader social imagination.

All epistemic injustice, and therefore testimonial injustice, is harmful to those who are on the receiving end. As Fricker (2007) explains, testimonial injustice causes the primary moral harm of being devalued or undermined as a knower. Insofar as being regarded as a knowing agent is a central part of human dignity and value, one's very humanity is undermined as well. In addition to this primary moral harm, Fricker identities a number of secondary harms that can result from testimonial injustice. These include epistemic harms such as a decline in one's epistemic self-trust (or a rise in one's proclivity toward doubting one's knowledge) and practical harms, which result from not being adequately listened to, heard, or believed. A possible (perhaps likely) practical harm in the case of the testimonial injustice against the innocent Black defendant, for example, is being wrongfully convicted. In all of these ways, testimonial injustice can be harmful to the speaker. But testimonial injustice also has wider social implications, which we want to draw attention to as well.

Importantly for our purposes is the following observation: where there is epistemic injustice, there is distrust in the capacity of members of socially marginalized groups to be knowers as a result of social stereotypes. Epistemic injustice—especially its testimonial form—is both caused by and worsens what we are calling "social distrust," which is distrust that occurs between different social groups. Specifically, there is a pervasive social distrust among dominant groups in the knowledge claims and testimonial assertions of members of marginalized groups. As a result, there is a corrupted epistemic sphere in which oppressed people are less able to transmit knowledge, and more privileged people have the domain of what they can know reduced substantially. The failure to be open to accepting the word of members of marginalized groups, and their reporting of their experiences, exacerbates the situated ignorance of more privileged people, where "situated ignorance" refers to gaps in one's knowledge that exist because of one's (more dominant) social position (see Dotson 2011). Such ignorance is worsened by the inability and/or unwillingness to take seriously the speech of marginalized people. The more ignorance there is about marginalized people and their experiences, the more distrust there is in and by them. There is social and epistemic distrust *in them* by others because they are viewed through the lens of harmful stereotypes, and there is social and epistemic distrust *by them* in more privileged groups because of the latter's situated ignorance.

In the following section, we will argue that algorithmic sorting and targeting on social media help to create the conditions that make epistemic injustice possible: the perpetuation and rigidity of stereotypes upon which epistemic injustice is based, the increased epistemic and moral distance between members of different social groups, the deepening of ignorance about the experiences of other social groups, and heightened social distrust.

5. Algorithms and Social Media, or The Perfect Storm for Epistemic Injustice

In his 2020 book *AI Ethics*, Mark Coeckelbergh argues that there is a need to pay serious attention to how algorithms are impacting our social interactions and moral lives. Importantly, he contends that these rapidly evolving technologies have the potential to "widen the gap between the powerful and powerless, thus accelerating injustice and inequality" (Coeckelbergh 2020, 75–76). We are interested in understanding how algorithmic processes on social media increase social injustice and oppression in our *interpersonal lives*. Our narrower goal is to show how algorithmic sorting and targeting can accelerate injustice and inequality *in the domain of knowledge.* We contend that algorithms on social media create the "perfect storm" for epistemic injustice because the very same conditions that make marginalized voices less likely to be heard and given proper uptake in general are compounded by the algorithmic processes of sorting and targeting, thereby creating a vicious cycle in which interpersonal epistemic injustices are worsened both on and offline. In what

follows, we will detail some of these conditions and how they manifest on social media.

Consider first how the algorithmic processes of sorting and targeting are likely to exacerbate harmful stereotypes (McIntosh 2016; Biddle 2019; Leetaru 2019a, 2019b; Barlas et al. 2020). People with stereotypical views can more freely express them online, receive uptake for them, or find communities of people who share and endorse them (R. Smith 2019, 15). To illustrate, consider that the most popular blogs written by women are the infamous "mommy blogs," which chronicle the ins and outs of "modern motherhood" and the "must-have" products needed to make things easier, cleaner, thinner, and smoother for other mothers. The lives of these "high-powered celebrity 'mommy-bloggers'" are not representative of most women's lives, nor of most *mothers'* lives. But the odds of women's actual lives making for a "successful" blog (or social media presence) "don't look so good, especially if [they] aim to write about something other than child-rearing, [their] rugged cowboy husband, or sponsored products" (Taylor 2014, 110). The domain of experiences likely to be shared on social media (the mechanism by which blogs gain exposure and popularity) is incredibly narrow.[15] Perhaps more perversely, it is steeped in traditional gendered norms.

Within the narrow epistemic worlds into which people are increasingly sorted, stereotypical representations of outsiders mostly go unchallenged because there are few voices of opposition or people with different perspectives or experiences to contribute. Stereotypical content (e.g., caricatures of people based on race, sexualized images of women, etc.) can also be explicitly targeted at people who are algorithmically determined to be sympathetic to it. The result of both this sorting and targeting is that people's stereotypical views become more rigid, which itself is a problem for epistemic justice.[16] As discussed above, stereotyping forms the basis of epistemic injustice; it causes those who are negatively stereotyped to have their epistemic status unfairly downgraded (see Fricker 2007).

Another feature of epistemic injustice that is closely tied to stereotyping is the creation and worsening of epistemic and moral distance between people of different identities, backgrounds, and belief systems. When such distance exists, it becomes less likely that people will effectively communicate across difference—that they will

---

[15] See Hindman (2008) for a more detailed analysis of this point. Hindman demonstrates that despite the fact that the Internet is widely heralded as a democratizing force, the public sphere is not actually made more inclusive by the Internet. Contrary to popular beliefs, Hindman argues, the Internet has done very little to broaden social discourse but rather empowers a small set of elites and already familiar perspectives.

[16] On the rigidity of stereotypes, see Blum (2004).

be able or willing to listen to or understand one another, to engage in meaningful dialogue or communication with another, or (as we will discuss below) to trust one another. Consider the following point from Zeynep Tufekci (2018), who notes that while social media technologies have the capacity to bring people together in new ways, they also have the potential to disperse—to create distance between—existing communities:

> Digital platforms allowed communities to gather and form in new ways, but they also dispersed existing communities, those that had watched the same TV news and read the same newspapers. Even living on the same street meant less when information was disseminated through algorithms designed to maximize revenue by keeping people glued to screens. It was a shift from a public, collective politics to a more private, scattered one, with political actors collecting more and more personal data to figure out how to push just the right buttons, person by person and out of sight.

Tufekci's observation is very much in line with our claim here: algorithmic processes on social media create, sustain, and exacerbate distance between people and groups. They do so, we contend, in at least two ways: by sorting users into different (and increasingly distant) online communities, and also by (further) entrenching social stereotypes, which further reinforce the apparent gaps between social groups. Essentially, the algorithms work in tandem with existing social stereotypes to divide people into different social and epistemic worlds. Stereotypes are divisive on their own because they contribute to "Othering," which occurs when certain groups are marked as deviant or inferior to dominant groups (see, for example, Beauvoir 2011; Young 1990). Othering further distances people and groups from one another: *morally* because they come to lack a sense of a "common humanity" (Blum 2004, 276) and *epistemically* because they are not positioned, or willing, to learn from one another (see Medina [2013] for a discussion on the epistemic vices of epistemic closed-mindedness and epistemic arrogance). Our claim is that algorithmic processes on social media exacerbate the social phenomenon of Othering, making people even more divided from one another.

To make the point more concrete, consider again the example of the "mommy bloggers" discussed above. Imagine a young mother—Anna—whose friend sends her a link to one of these mommy blogs through Facebook. It's an entertaining read so Anna "likes" it. Next thing she knows, she gets a targeted ad for the Roomba automated vacuum that the blogger said was a "godsend" with her messy toddlers. Then, Anna gets a targeted ad for the high-end concealer that was sponsored in the blog as well—a concealer meant to hide those inevitable dark circles under an

exhausted new mother's eyes. She glances over to her mirror. She does look tired after all. So, she clicks on the ad link and places an order for the concealer. Facebook's algorithm learns that the target ad worked. Before long, an increasing number of similar blogs, related posts, and motherhood-related memes are being recommended to Anna by Facebook's algorithms. Gradually, she is presented with an increasingly homogeneous image of motherhood—one that is white, straight, thin, and moneyed. This narrow image of motherhood becomes reinforced for her. Over time, other images of motherhood and parenting start to feel increasingly "off" to her—perhaps they don't register as "real" motherhood at all. Mothers and parents who do not fit the bill become "Other"—lesser or deviant. At the same time, those who are not reflected in these images of motherhood—mothers and parents who do not fit the narrow, stereotypical, and normative presentation of motherhood—might understandably come to resent it (and, moreover, to resent those whose experiences are consistent with this status quo). In both regards, there is a growing distance between these different groups of "mothers" and how they see and understand one another, if they see or understand one another at all.

Growing distance between groups (and related gaps in awareness and understanding of each other's experiences) is at the core of pernicious social and epistemic ignorance (see for example Mills 2007; Dotson 2011). Over time, a lack of exposure to the lives, perspectives, and experiences of others deepens our extant ignorance and widens the gaps in our understanding (see Dotson 2011). Moreover, when one sees their own experiences and worldviews disproportionately reflected in the content presented to them, it makes it harder to recognize and appreciate the limits of one's own experience and knowledge. In other words, it makes it more difficult to recognize and appreciate that one is both *socially and epistemically situated* and that one's perspectives, beliefs, and knowledge are necessarily partial and limited as a result. Such ignorance—both the ignorance of others' lives and experiences *and* the ignorance of our own epistemic limitations—is at the heart of epistemic injustice. We are less likely to listen to, trust, and believe others when we are fundamentally unaware of their experiences. Worse yet, this generates a vicious cycle, because when we disbelieve or distrust others, the gaps in our knowledge cannot close—the ignorance is thereby reinforced. There is a tight and pernicious connection between social and epistemic ignorance and epistemic injustice.

Our claim is that algorithmic processes on social media speed up this vicious cycle by widening the gaps that drive ignorance and, as a result, worsening the problem of epistemic injustice.[17] This is because the algorithmic processes, as discussed above, sort out different perspectives and experiences—those of anyone

---

[17] For a discussion of ignorance as it relates to technology, see Greyson (2019). For a discussion of challenges to online ignorance see Frost-Arnold (2016).

deemed "Other." Our epistemic worlds increasingly reflect the experiences of people like us. For the socially dominant, this is particularly problematic, as nondominant voices become shielded from view, allowing dominant narratives to remain unshaken. To revisit the example of Anna, a Facebook feed that presents a user with a narrow and privileged image of motherhood does not allow one to learn about or come to understand the lived experiences of, for example, queer mothers, low-income and working-class mothers, or mothers who experience racial discrimination against themselves and their children. This allows biased and stereotypical assumptions about what "real" motherhood looks like to remain virtually unchallenged.

Lastly, we contend that algorithmic processes on social media can also (further) undermine social and epistemic trust. Algorithmic sorting and targeting lead people to inhabit epistemic worlds largely inhabited by others who think like they do and with whom they come to form "in-group" trust (Nguyen 2020) and a general sense of belonging. By contrast, people outside that epistemic world are distrusted. The successful acquisition and transfer of knowledge requires trust (Hardwig 1991). Trust, in other words, is a precondition for the ability to communicate effectively with others, especially across social difference, vulnerability, or imbalances in power. Insofar as algorithms on social media have the effect of diminishing social and epistemic trust, they (further) impede the possibility of meaningful communication and knowledge exchange across social difference (e.g., between people who find themselves in distant epistemic and social worlds). When this impediment negatively impacts members of marginalized groups, so that they become increasingly unable to secure epistemic uptake, epistemic injustice occurs.

Without describing it as "epistemic injustice," Avriel Epps-Darling (2020) raises a similar concern—namely, that the effect of algorithms on social media are likely to render people less able to understand one another in an increasingly diverse social world. She writes:

> Living in a world controlled by discriminatory algorithms can further segregate white youths from their peers of color. TikTok's content-filtering algorithm, for example, can drive adolescents toward echo chambers where everyone looks the same. This risks diminishing teens' capacity for empathy and depriving them of opportunities to develop the skills and experiences necessary to thrive in a country that's growing only more diverse.

She adds:

> Algorithmic racism exists in a thriving ecosystem of online discrimination, and algorithms have been shown to amplify the voices

of human racists. Black teens experience an average of five or more instances of racism daily, much of it happening online and therefore mediated by algorithms. Radicalization pipelines on social platforms such as YouTube can lead users down rabbit holes of videos designed to recruit young people, radicalize them, and inspire them to commit real-world violence. Before the internet, parents could discourage their kids from spending time with bad influences by monitoring their whereabouts. Today, teens can fraternize with neo-Nazis and spread eugenics propaganda while just feet away from a well-intentioned but unaware parent. Part of the problem is that parents don't see the underlying structures of popular platforms, such as YouTube, Facebook, and Reddit, as strangers that can take the hand of a teenager and guide them deeper and deeper into disturbing corners of the web. There is no "stranger danger" equivalent for a recommendation engine. (Epps-Darling 2020)

What Epps-Darling is describing is also what concerns us here: algorithmic sorting and targeting impacting what individual users see, and do not see, in their social media feeds (see also Stern 2021). As a result, they shape people's informational and social landscapes substantially. This distorting of how we see and relate to one another is likely to worsen our ability to trust others and to fairly evaluate or give uptake to the knowledge claims of those who are unlike us. When this happens as a result of, for example, the targeting of misogynistic or white supremacist content to users, the consequence is a disproportionate impact on those who already experience epistemic marginalization.

Readers might be wondering what can be done to offset this problem. Can't the social media sites just use the power of algorithms to improve the epistemic landscape by rooting out biased, discriminatory, or stereotypical content targeting marginalized groups? In theory, yes. But so far, this isn't proving to be a promising line of defense. So far, efforts to restrict hate speech targeted at members of marginalized groups are backfiring, instead disproportionately flagging the speech of marginalized people. As Shirin Ghaffary (2019) explains, platforms such as Facebook and Twitter are using algorithms in efforts to curb hateful speech on their sites. The idea is that the algorithms will be better equipped to identify this speech and do so quickly. The motivation and urgency for these platforms to step up their efforts to reduce hateful speech on their platform is largely in response to recent mass shootings and violence that have been linked to hate speech online, targeted at marginalized communities (see Hatzipanagos 2018). However, multiple studies have found that the algorithms trained to identify hate speech end up amplifying racial bias. One study indicated that the algorithm used to detect hate speech was one and

a half times more likely to flag tweets written by Black users as offensive, and even more so when the tweet was written in Black vernacular (Sap et al. 2019). This overrepresentation of Black users' tweets in efforts to detect hateful speech has been reflected in other studies as well (e.g., Davidson, Bhattacharya, and Weber 2019).[18] If the algorithms designed to remove such content disproportionately flag the speech of users from marginalized groups, then they *heighten* epistemic injustice rather than combat it.[19]

It is important to note that social media executives, such as Facebook's Mark Zuckerberg, are well aware that their platforms are turning people against each other

---

[18] One issue is that speech, and the meaning of speech, can be incredibly context dependent. For example, uses of the "N-word" or terms such as "queer" or "dyke" can constitute slurs, or otherwise be offensive, in some contexts but not in all contexts. The algorithms currently employed by platforms such as Facebook and Twitter cannot account for this context-specificity. So, for example, when an algorithm is directed to flag or remove tweets that use the "N-word" or "queer," there is no ability to adjudicate between offensive and inoffensive uses, or in-group vs. out-group uses (see Anderson 2018). Bias is baked in from the start. The result is that the inoffensive speech of, for example, Black or LGBTQ+ users is disproportionately flagged or removed, while legitimately offensive speech is disproportionately left unscathed.

Some argue that the problem goes even deeper than this and that the speech of Black users is itself more heavily policed than that of white users. For example, in one notable instance, a Black user was banned from Facebook for sharing a "Dear White People" note that many of her white friends had posted without consequence (Sankin 2017). Review of internal documents from Facebook suggest that white men are indeed more insulated from having their speech flagged and removed (see Angwin 2017). The disproportionate silencing of Black (and other marginalized) users is itself reflective of epistemic injustice, and something that demands our attention.

[19] Another possible (albeit partial) solution is found at the level of individual users: individual users can intentionally curate their social media networks so as to ensure greater levels of exposure to a plurality of different (and underrepresented) voices—e.g., by being intentional about who they add or follow and by following more people with different perspectives. This would, of course, be a limited solution, as the algorithms would still influence what content was highlighted and "driven up." Moreover, having the foresight to be intentional in this way would require there to be far more social understanding of the influence of algorithms on social media, as well as a broader concern about the epistemic effects of social media (e.g., that social media networks are just as much epistemic and informational networks as social ones). Both requirements seem fairly implausible at present.

and having corrosive social and epistemic effects (Hatmaker 2021). And yet, they refuse to do anything to stop it (A. Smith 2020; Frier 2021). Part of the uphill battle of ending algorithmic-driven injustice (epistemic and otherwise) is getting the platforms to choose justice and what is right over a constant drive toward increasing profit or getting governments to engage in meaningful oversight (Swisher 2020). Both, unfortunately, seem like a long shot.

6. On Algorithmic Injustice and Oppression, and the Need to Think Broadly

The issue we have identified—namely, the contribution that algorithms make to epistemic injustice—is but one dimension of a much broader problem generated by the algorithms rapidly infiltrating our lives. Some have named the problem "algorithmic injustice" or "algorithmic oppression," and we believe there is room to flesh out both of these concepts more comprehensively. Here, we want to press on each of them and suggest that a robust accounting of the injustice and oppression that algorithms cause must also include the seemingly subtle ways in which algorithms influence our social worlds, including, as we have detailed here, our epistemic interactions and associated trust relations.

One leader in the movement to end the perverse social implications of rapidly developing AI is MIT's Joy Buolamwini (see Algorithmic Justice League, n.d.). Buolamwini brought significant attention to the issue of algorithmic bias with a 2016 TED talk, which now has 1.4 million views and counting (Buolamwini 2016).[20] Buolamwini's work on issues of algorithmic bias and injustice began with her discovery that facial recognition technologies are both racist and sexist.[21] Now, her work is to uncover the many ways in which AI systems perpetuate racism, sexism, ableism and other harmful forms of discrimination. The Algorithmic Justice League (AJL), which Buolamwini founded, is aimed at illuminating these social implications of AI and advocating for more justice within the domain of digital technology.

The work of the AJL is invaluable, and Joy Buolamwini has been a vital voice within the movement for accountability for the harms perpetuated by AI systems. However, it is worth noting that much of the focus of AJL research and advocacy focuses on injustices at the "macro" scale and most often those being perpetuated by institutions such as medical institutions, the criminal justice system, banking and crediting agencies, and big tech (i.e., through the use of technologies like facial recognition). Issues pertaining to racial profiling, surveillance, housing or employment discrimination, and the like are incredibly important—we certainly do not want to

---

[20] Also see Shalini Kantayya's (2020) documentary, *Coded Bias*, which features Buolamwini.

[21] They are disproportionately able to consistently and accurately identify white male faces.

deny or minimize the significance of these types of discrimination—but they are only part of the story when it comes to the sorts of injustice that algorithms can (and do) perpetuate. Algorithms can also contribute to injustices that are more "micro" in scale. They are less overt or obvious, they occur at an interpersonal level, and they are often not quantifiable in the way that, for example, a denied rental application or an unfairly evaluated job application is quantifiable. They are also more "everyday" sorts of injustices that include epistemic injustice.[22] We think it is important to pay attention to them.

What we have in mind is described quite clearly by Avriel Epps-Darling (2020), who notes that "acts of technological racism might not always be blatant," likening the modality of "everyday" algorithmic harms to those of another phenomenon, microaggressions.[23] She writes:

> Algorithmic racism frequently functions as a type of technological microaggression—those thinly veiled, prejudiced behaviors that often happen without the aggressor intending to hurt anyone. But the algorithmic variety differs from human microaggressions in several ways. For one, a person's intent might be hard to pin down, but the computational models imbued with algorithmic bias can be exponentially more opaque. Several common machine-learning models, such as neural networks, are so complex that even the engineers who design them struggle to explain precisely how they work. Further, the frequency at which technological microaggressions occur is potentially much higher than in real life because of how much time teens spend on devices, as well as the automatic, repetitive nature of programmed systems. And everyone knows that human opinions are subjective, but algorithms operate under the guise of

---

[22] Relevant here is Iris Marion Young's claim that while some manifestations of oppression are overt, oppression also refers to the "disadvantage and injustice some people suffer . . . because of the everyday practices of a well-intentioned liberal society" (Young 1990, 39). Its causes, Young says, "are embedded in unquestioned norms, habits, and symbols, in the assumptions underlying institutional rules and the collective consequences of following those rules" (39).

[23] Microaggressions are widely understood to be "verbal, nonverbal, and/or environmental slights, snubs, or insults that are either intentional or (most often) unintentional; they convey hostile, derogatory, or otherwise negative messages to target persons based upon their membership in a structurally oppressed social group" (Freeman and Stewart 2018, 412; citing Sue 2010).

> computational objectivity, which obscures their existence and lends legitimacy to their use. (Epps-Darling 2020)

We believe the Algorithmic Justice League would be open to broadening their understanding of algorithmic injustice to encompass thinly veiled, everyday injustices, including those of an epistemic sort. As they note on their website:

> The deeper we dig, the more remnants of prejudice we will find in our technology. We cannot afford to look away this time because the stakes are simply too high. We risk losing the gains made with the civil rights movement and other movements for equality under the false assumption of machine neutrality. (Algorithmic Justice League, n.d.)

We cannot agree more. Now that significant amounts of our lives are spent online, we need to uncover both overt and more subtle algorithmic injustices. We need a fine-tooth comb to expose the latter and ensure they are not ignored. Our hope is that we have aided in this effort by revealing one such injustice: epistemic, and especially testimonial, injustice.

A related concept to algorithmic injustice is "algorithmic oppression," which can be largely tied to the research of a UCLA associate professor of information studies, Safiya Umoja Noble. In her 2018 book, *Algorithms of Oppression,* Noble examines the ways in which the algorithms are biased and perpetuate harmful stereotypes on the basis of race and gender.[24] Although Noble's focus is more narrowly about the subtle ways in which *search engines* reinforce oppression, the concept of algorithmic oppression ought to be thought of more broadly: to apply to the myriad ways in which algorithms reproduce oppressive structures. It seems as if Noble was certainly hoping that her work would be extended in these ways. As she writes in her conclusion, "This book can open up a dialogue about radical interventions on socio-technical systems in a more thoughtful way that does not further marginalize people who are already in the margins. Algorithms are, and will continue to be, contextually relevant and loaded with power" (Noble 2018, 171).

In short, we believe that broad conceptualizations of algorithmic injustice and oppression are needed, ones that account for the many ways (both "macro" and "micro") in which algorithms can cause injustice and oppression. Such accounts are currently lacking in the literature. We hope that those who are engaging in critical work about the social implications of digital technologies will continue to uncover and

---

[24] For a discussion of algorithms in search engines, and problematic "autocomplete results," see Sumpter (2018, 188–94); also see Robert Smith (2019, x) and Miller and Record (2017).

analyze different dimensions of algorithmic injustice and oppression and add clarity and concision to these concepts. Our contribution here has been to highlight a specific manifestation of these problems—namely, how algorithms on social media contribute to epistemic injustice or oppression within the domain of knowledge and do so often in subtle ways.

7. Conclusion

One overlooked risk of the influence of social media on our lives is that it worsens the problem of epistemic injustice and related problems of diminished social and epistemic trust and poor communication across social difference. We have identified two related processes that contribute to these effects: what we've called *algorithmic sorting* and *algorithmic targeting.* The former creates greater distance between, and lessens exposure to, differing perspectives. It thus contributes to an Othering of those unlike us. The latter is even more insidious; problematic content that drives disdain for certain groups gets targeted directly to people who might already have (or might be likely to form) biases against those out-groups. Targeted messages about members of marginalized groups, then, serve as confirmatory evidence for one's preexisting biases, which algorithms have been able to detect on the basis of like patterns and content-viewing histories. Together, algorithmic sorting and targeting on social media create the perfect conditions for epistemic injustice—conditions in which outsiders are viewed as Other, where there is decreasing exposure to and engagement with different perspectives, where different perspectives are less likely to be trusted or seen as reliable, and where biases and stereotypes are reinforced via an endless flow of what appears to be confirmatory evidence.

Epistemic injustice, as well as the harms it causes for speakers and the degrading effect it has on our ability to place an appropriate degree of trust in the word of others, is a serious moral and social problem. It is a problem that compounds other injustices that members of structurally oppressed groups face, as it makes them less able to speak out effectively regarding such injustice. It deepens social division, as it renders us less able to communicate with and place trust in one another across social difference. To the extent that we are beginning to sharpen our focus on the role that algorithms play in reinforcing injustice and oppression, we must pay careful attention to the harmful effects that such AI has in the already imperfect epistemic domain.

References

Algorithmic Justice League. n.d. "About." Accessed May 2021. https://www.ajl.org /about.

Anderson, Luvell. 2018. "Calling, Addressing, and Appropriation." *Bad Words: Philosophical Perspectives on Slurs,* edited by David Sosa, 6–28. Oxford: Oxford University Press.

Angwin, Julia. 2017. "Facebook's Secret Censorship Rules Protect White Men from Hate Speech but Not Black Children." *ProPublica,* June 28, 2017. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.

Barlas, Pinar, Kyriakos Kyriakou, Olivia Guest, Styliani Kleanthous, and Jahna Otterbacher. 2020. "To 'See' Is to Stereotype: Image Tagging Algorithms, Gender Recognition, and the Accuracy-Fairness Trade-Off." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW3): 1–31. https://doi.org/10.1145/3432931.

Beauvoir, Simone de. 2011. *The Second Sex*. Translated by Constance Borde and Sheila Malovany-Chevallier. New York: Vintage.

Biddle, Sam. 2019. "Facebook's Ad Algorithm Is a Race and Gender Stereotyping Machine, New Study Suggests." *Intercept,* April 3, 2019. https://theintercept.com/2019/04/03/facebook-ad-algorithm-race-gender/.

Blum, Lawrence. 2004. "Stereotypes and Stereotyping: A Moral Analysis." *Philosophical Papers* 33 (3): 251–89. https://doi.org/10.1080/05568640409485143.

Bruns, Axel. 2019. *Are Filter Bubbles Real?* Cambridge: Polity Press.

Buolamwini, Joy. 2016. "How I'm Fighting Bias in Algorithms." Filmed November 2016 in Brookline, MA. TEDx video, 8:35. https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms.

Campbell, Hebe, and Matthew Holroyd. 2021. "Instagram Algorithms Push Misinformation to Users, Study Claims." *Euronews*, updated March 9, 2021. https://www.euronews.com/2021/03/09/instagram-algorithms-push-misinformation-to-users-study-claims.

Coeckelbergh, Mark. 2020. *AI Ethics.* Cambridge, MA: MIT Press.

Collins, Patricia Hill. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, 2nd ed. New York: Routledge.

Davidson, Thomas, Debasmita Bhattacharya, and Ingmar Weber. 2019. "Racial Bias in Hate Speech and Abusive Language Detection Datasets." In *Proceedings of the Third Workshop on Abusive Language Online*, 25–35. Florence, Italy: Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3504.

Dotson, Kirstie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–57. https://doi.org/10.1111/j.1527-2001.2011.01177.x.

Epps-Darling, Avriel. 2020. "How the Racism Baked Into Technology Hurts Teens." *Atlantic*, October 24, 2020.https://www.theatlantic.com/family/archive/2020

/10/algorithmic-bias-especially-dangerous-teens/616793/.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Freeman, Lauren, and Heather Stewart. 2018. "Microaggressions in Clinical Medicine." *Kennedy Institute of Ethics Journal* 28 (4): 411–49. https://doi.org /10.1353/ken.2018.0024.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Frier, Sarah. 2021. "The 'Ugly Truth' about Facebook." Review of *An Ugly Truth: Inside Facebook's Battle for Domination*, by Sheera Frenkel and Cecilia Kang. *New York Times*, July 9, 2021. https://www.nytimes.com/2021/07/09/books /review/the-ugly-truth-sheera-frenkel-and-cecilia-kang.html.

Frost-Arnold, Karen. 2016. "Social Media, Trust, and the Epistemology of Prejudice." *Social Epistemology* 30 (5-6): 513–31. https://doi.org/10.1080/02691728 .2016.1213326.

Ghaffary, Shirin. 2019. "The Algorithms that Detect Hate Speech Online Are Biased against Black People." *Vox*, August 15, 2019. https://www.vox.com/recode /2019/8/15/20806384/social-media-hate-speech-bias-black-african-american- facebook-twitter.

Greyson, Devon. 2019. "The Social Informatics of Ignorance." *Journal of the Association for Information Science and Technology* 70 (4): 412–15. https://doi.org/10.1002/asi.24143.

Hardwig, John. 1991. "The Role of Trust in Knowledge." *Journal of Philosophy* 88 (12): 693–708. https://doi.org/10.2307/2027007.

Hatmaker, Taylor. 2021. "Facebook Knows Instagram Harms Teens. Now Its Plan to Open the App to Kids Looks Worse than Ever." *Tech Crunch*, September 16. https://techcrunch.com/2021/09/16/facebook-instagram-for-kids-mosseri- wsj-teen-girls/.

Hatzipanagos, Rachel. 2018. "How Online Hate Turns into Real-Life Violence." *Washington Post*, November 30. https://www.washingtonpost.com/nation /2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/.

Hindman, Matthew. 2008. *The Myth of Digital Democracy*. Princeton, NJ: Princeton University Press.

Jennings, Rebecca. 2018. "Why Targeted Ads Are the Most Brutal Blows." *Vox*, September 25, 2018. https://www.vox.com/the-goods/2018/9/25/17887796 /facebook-ad-targeted-algorithm.

Kantayya, Shalini, dir. 2020. *Coded Bias*. 7th Empire Media. Released to internet November 2020 on Netflix streaming video, 85 min. https://www.netflix.com /title/81328723.

Kantrowitz, Alex. 2020. "Discussing the Future of Social Media–Driven Protests with Zeynep Tufekci." *One Zero*, August 24, 2020. https://onezero.medium.com /discussing-the-future-of-social-media-driven-protests-with-zeynep-tufekci-a5c914c34176.

Kidd, Ian James, José Medina, and Gaile Pohlhaus Jr. 2017. *The Routledge Handbook of Epistemic Injustice*. New York: Routledge. https://doi.org/10.4324/9781315 212043.

Leetaru, Kalev. 2019a. "Do Instagram's Ads Reinforce Harmful Stereotypes?" *Forbes*, June 19, 2019. https://www.forbes.com/sites/kalevleetaru/2019/06/19/do-instagrams-ads-reinforce-harmful-stereotypes/?sh=7173c171c488.

———. 2019b. "Is It Time to Regulate Biased and Harmful Stereotypes in Social Media Ads?" *Forbes*, June 21, 2019. https://www.forbes.com/sites/kalevleetaru /2019/06/21/is-it-time-to-regulate-biased-and-harmful-stereotypes-in-social-media-ads/?sh=34f4924129a3.

Long, Katherine Anne. 2021. "Amazon's Algorithms Promote Extremist Misinformation, Report Says." *Seattle Times*, April 28, 2021. https://www.seattletimes.com/business/amazon/amazons-algorithms-promote-extremist-misinformation-report-says/.

Lorde, Audre. 1984. *Sister Outsider: Essays and Speeches*. Trumansburg, NY: Crossing Press.

McIntosh, Bennett. 2016. "Bias in the Machine: Internet Algorithms Reinforce Harmful Stereotypes." Princeton University, Department of Computer Science website, November 22, 2016. https://www.cs.princeton.edu/news/bias-machine-internet-algorithms-reinforce-harmful-stereotypes.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press.

Miller, Boaz, and Isaac Record. 2017. "Responsible Epistemic Technologies: A Social-Epistemological Analysis of Autocompleted Web Search." *New Media & Society* 19 (12): 1945–63. https://doi.org/10.1177/1461444816644805.

Mills, Charles W. 2007. "White Ignorance." In *Race and Epistemologies of Ignorance*, edited by Shannon Sullivan and Nancy Tuana, 11–38. Albany: State University of New York Press.

Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17 (2): 141–61. https://doi.org/10.1017/epi.2018.32.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press.

Paul, Kari. 2021. "'It Let White Supremacists Organize': The Toxic Legacy of Facebook's Groups." *Guardian*, February 4, 2021. https://www.theguardian.com/tech nology/2021/feb/04/facebook-groups-misinformation.

Pierce, Katie. 2019. "Zeynep Tufekci on Tech's Powers and Perils for Democracy." Johns Hopkins University HUB, February 11, 2019. https://hub.jhu.edu/2019 /02/11/zeynep-tufekci-democracy-dialogues/.

Polson, Nicholas G., and James Scott. 2018. *AIQ: How People and Machines Are Smarter Together*. New York: St. Martin's Press.

Rubel, Alan, Clinton Castro, and Adam K. Pham. 2021. *Algorithms and Autonomy: The Ethics of Automated Decision Systems*. Cambridge: Cambridge University Press.

Sankin, Aaron. 2017. "How Activists of Color Lose Battles against Facebook's Moderator Army." *Reveal News*, August 17, 2017. https://revealnews.org /article/how-activists-of-color-lose-battles-against-facebooks-moderator-army/.

Sap, Maarten, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. "The Risk of Racial Bias in Hate Speech Detection." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–78. Florence, Italy: Association for Computational Linguistics. https://doi.org/10.18653/v1 /P19-1163.

Smith, Adam. 2020. "Facebook Knew Its Algorithm Made People Turn Against Each Other but Stopped Research." *Independent*, May 28, 2020. https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook -algorithm-bias-right-wing-feed-a9536396.html.

Smith, Robert Elliot. 2019. *Rage inside the Machine: The Prejudice of Algorithms, and How to Stop the Internet Making Bigots of Us All*. London: Bloomsbury Business.

Stephen, Bijan. 2015. "Get Up, Stand Up: Social Media Helps Black Lives Matter Fight the Power." *Wired*, November 2015, 120–24. https://www.wired.com/2015 /10/how-black-lives-matter-uses-social-media-to-fight-the-power/.

Stern, Joanna. 2021. "Social-Media Algorithms Rule How We See the World: Good Luck Trying to Stop Them." *Wall Street Journal*, January 17, 2021. https://www.wsj.com/articles/social-media-algorithms-rule-how-we-see-the-world-good-luck-trying-to-stop-them-11610884800.

Sue, Derald Wing. 2010. *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation*. Hoboken, NJ: John Wiley & Sons.

Sumpter, David. 2018. *Outnumbered: From Facebook and Google to Fake News and Filter-Bubbles—The Algorithms That Control Our Lives*. New York: Bloomsbury Sigma.

Swisher, Kara. 2020. "When It Comes to Facebook, the Need for Action Has Been Obvious for a Long Time." *New York Times*, December 10, 2020. https://www.nytimes.com/2020/12/10/opinion/facebook-antitrust-ftc.html.

Taylor, Astra. 2014. *The People's Platform: Taking Back Power and Culture in the Digital Age*. Toronto: Random House Canada.

Tufekci, Zeynep. 2014. "Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape." *Journal of International Affairs* (New York) 68 (1): 1–18.

———. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest.* New Haven, CT: Yale University Press.

———. 2018. "How Social Media Took Us from Tahrir Square to Donald Trump." *MIT Technology Review*, August 14, 2018. https://www.technologyreview.com /2018/08/14/240325/how-social-media-took-us-from-tahrir-square-to-donald-trump/.

Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.

HEATHER STEWART is assistant professor of philosophy at Oklahoma State University. Her primary areas of research are within bioethics and philosophy of medicine, feminist philosophy, and social and political philosophy. Specifically, she works on issues of speech and language and the ways in which our communicative practices reflect and reinforce unjust social hierarchy and systems of oppression. Her work has appeared in several journals including the *Kennedy Institute of Ethics Journal, Perspectives on Psychological Science,* and *Southwest Philosophy Review*. She is currently cowriting a book manuscript titled *Microaggressions in Medicine*, which is under contract with Oxford University Press.

EMILY CICHOCKI is a PhD candidate in philosophy at Western University. Her research is primarily in the areas of social and political philosophy, feminist philosophy, and moral psychology. She is currently working on a dissertation that focuses on the concepts of structural injustice, complicity, and solidarity. Part of her dissertation examines the effects that algorithms have on solidarity in the digital age. She is also part of an interdisciplinary team dedicated to research at the intersection of AI ethics and social justice. In addition to her research, Emily is passionately involved in labor activism.

CAROLYN McLEOD, FRSC, is professor and chair of philosophy at Western University. Author of *Self-Trust and Reproductive Autonomy* (MIT) and *Conscience in Reproductive Health Care* (Oxford), she is an expert on ethical issues in reproductive

health care. She has also done extensive research on the ethics of adoption and on key concepts in moral philosophy such as trust and autonomy. Her current work focuses on the distrust of socially marginalized groups in public institutions, particularly those of health care and child welfare. She is concerned about the increased use of AI by these institutions and its likely impact on the institutional distrust of marginalized people.