JUST TECH

FIELD REVIEW

Impact Assessment of Human-Algorithm Feedback Loops

By: Nathan Matias, Lucas Wright

Matias, Nathan and Lucas Wright. “Impact Assessment of Human-Algorithm Feedback Loops.” Just Tech. Social Science Research Council. March 1, 2022. DOI: https://doi.org/10.35650/JT.3028.d.2022

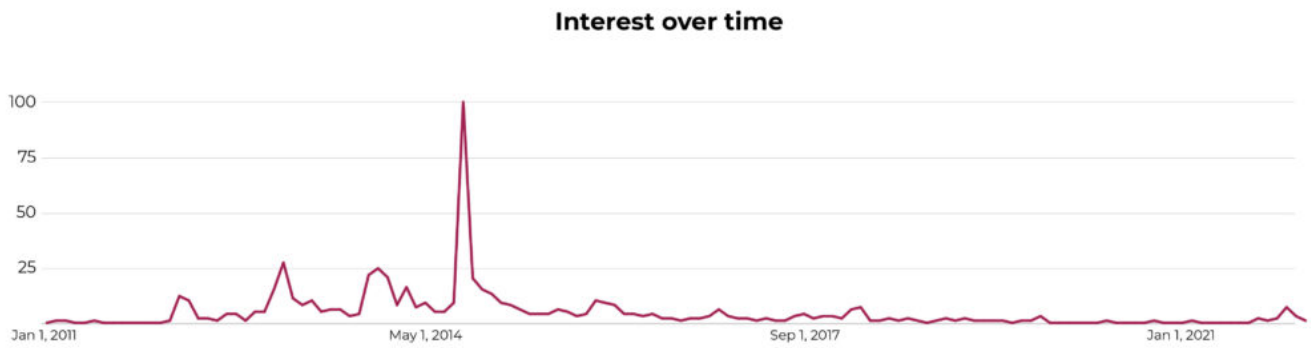
ABSTRACT

How can we understand and manage the impact of adaptive algorithms that respond to people’s behavior and also influence what people do? From predictive policing to video recommendations, these algorithms shape outcomes including criminal justice, economic inequality, public health, and social change. This field review provides an introduction to the challenges of governing adaptive algorithms, whose interactions with human behavior cannot yet be reliably predicted or assessed. This review is a practical tool for regulators, advocates, journalists, scientists, and engineers who are working to assess the impact of these algorithms for social justice.

The review begins by describing areas of social justice impacted by adaptive algorithms. It then describes human-algorithm feedback, names common feedback patterns, and links those patterns to long-standing injustices and opportunities for social change. It outlines fundamental advances that are still needed for effective impact assessment, including new forms of knowledge, new interventions for change, and governance that involves affected communities. The review concludes with high-level recommendations for anyone working to assess the impact of human-algorithm feedback.

In March 2017, when Wilmer Catalan-Ramirez was tackled in his Chicago home by immigration agents, threatened with deportation, and detained, he couldn’t understand why. The agents had no warrant. He had no criminal record, and his only encounters with police had been random stops in his neighborhood. Unknown to Catalan-Ramirez, his name had been added to a Chicago predictive policing system after he was injured in a neighborhood drive-by shooting (Moreno 2017). This Strategic Subjects List, according to police, was designed to improve community support. In practice, police targeted people and communities on the list, trapping them in a cycle of escalating suspicion (Asher and Arthur 2017).

Actor Jennifer Lawrence was caught in another escalating cycle in 2014 when her intimate photos were stolen and circulated on the social platform Reddit. As people clicked and up-voted the pictures, Reddit's algorithms showed them to even more people, encouraging the algorithm further. When asked about the experience, Lawrence told *Vanity Fair*, "It just makes me feel like a piece of meat that's being passed around for a profit" (Kashner 2014). Reddit made enough money from the incident to fund its computer systems for a month (Greenberg 2014).



Google Searches for "Jennifer Lawrence" on the Google search engine reached its highest all-time value in 2014 as people looked for her stolen intimate photos.

Could the harms experienced by Catalan-Ramirez and Lawrence have been predicted or prevented? Advocates have called for algorithm impact assessments to help manage the risks from algorithms. But unlike detection, risk-assessment, and decision-making systems that are designed for accuracy and consistency, these algorithms are designed to change over time. When algorithms adapt, their behavior is hard to predict, fairness is usually unattainable, and effective remedies are even harder to identify—for now.

Catalan-Ramirez and Lawrence were both victims of feedback between human and algorithm behavior. Feedback happens when humans and algorithms react to each other in ways that change algorithms without further involvement from engineers. And this feedback is everywhere—directing law enforcement, managing financial systems, shaping our cultures, and flipping a coin on the success or failure of movements for change. Since human-algorithm feedback is already a basic pattern in society, we urgently need ways to assess the impact of ideas for steering those patterns toward justice.

A Social Justice View of Human-Algorithm Behavior

Conversations about social justice and algorithms often focus on the bias and accuracy of decisions by algorithms. Bias happens when an algorithm for facial recognition, court sentencing, or credit scoring tends to mistake or wrongly penalize a group of people on average. Bias is a useful concept for evaluating judgments that need to be impartial and independent (Barocas, Hardt, and Narayanan 2019).

However, evaluations of bias could not have protected Catalan-Ramirez or Lawrence because adaptive algorithms cannot be tested for bias—they work like a mirror rather than an independent, impartial judge. The Reddit ranking system was designed to reflect the most popular content rather than provide

an unbiased judgment of each picture's social value. Chicago's Strategic Subjects List was a popularity ranking for policing. The system was deliberately designed to direct police toward people who already had more contact with public services and the police. Since bias applies to impartial judges but not to mirrors, we need different language to recognize the problems of adaptive algorithms. We also need different forms of power to solve those problems.

Bias reduction is too small a vision in a world where injustice is the norm. Society would not be better off if Reddit equally amplified stolen intimate pictures of famous men or if Chicago also overpoliced white neighborhoods. Instead, we need to imagine a positive vision of a just society and work toward that goal (Costanza-Chock 2020).

Photo by Magnus Mueller

What do we mean by social justice? In a persistently unequal society built with stolen land, stolen labor, and stolen opportunity, social justice involves economic justice (McGhee 2021). Criminal and border justice require reforms to a system that has made African Americans second-class citizens in their own country (Alexander 2000) and has created an underclass from the undocumented Americans who power our economy (Chavez 2012). Health equity addresses a history of mistreatment and underprovision of health resources to marginalized communities (Donohoe 2012; Reverby 2012).

A positive vision of social justice also includes the institutions of democracy and collective social power. The right to vote is a cornerstone of justice in the United States, where power holders have continually suppressed voting rights (Anderson 2018). Beyond voting, collective behavior arises from how people think and act. That's why epistemic justice—what people know and believe about others through communication and media—also forms a basic building block of social justice (Fricker 2007).

As adaptive algorithms orchestrate more of our economy, criminal justice systems, healthcare, and democratic institutions, a more equitable and just future depends on our ability to understand and change human-algorithm feedback.

Understanding Human-Algorithm Feedback

Both Chicago's predictive policing system and the Reddit algorithm were designed to adapt to the world around them, observing changing situations and adjusting their behavior in turn. Feedback happens when adaptive algorithms react to the behavior of humans, who are also reacting to the algorithms.

Investors learned about feedback in 2010, when they lost billions of dollars in just a few minutes. In high-frequency stock market trading, investors use algorithms to buy and sell thousands of investments per day. When humans start selling investments rapidly, algorithms can react by selling stocks, too, spiraling rapidly into a stock market "flash crash" (Virgilio 2019).

Chicago's predictive policing system created a similar spiral—but one created by injustice. With each police patrol, the software was fed new information about who officers met. The software then updated the list to prioritize people who had more contact with the police. When police followed those recommendations, the feedback increased policing for communities that were already overpoliced (Saunders, Hunt, and Hollywood 2016). In the stock market and on the streets of Chicago, harmful outcomes were caused by a cycle of influence between humans and algorithms.

Feedback can come from one person or a million people. When you use a predictive keyboard on your phone, your choice of words is shaped by your keyboard's suggestions. In turn, this personalized software adapts its suggestions to the messages you write, making your language less creative (Arnold, Chauncey, and Gajos 2020). Other systems, like Reddit's rankings, aggregate behavior across many people (Ekstrand, Riedl, and Konstan 2011). Some systems do both (Li et al. 2010).

Adaptive algorithms do not create injustice on their own, but they do amplify it when they respond to unjust laws, institutions, or human behavior.

Adaptive algorithms do not create injustice on their own, but they do amplify it when they respond to unjust laws, institutions, or human behavior. Reddit did not invent a sexist culture of voyeuristic abuse

when its aggregator encouraged people to view Lawrence's pictures without permission. Celebrity gossip websites also published the images, which were also distributed on file-sharing servers (Sparkes 2014). But in a world where sexism is already powerful and profitable, Reddit's algorithm poured gasoline onto the open flame of misogyny.

Because feedback amplifies collective behavior, it can also grow collective power for social change (Yasseri et al. 2016). In the summer of 2020, when Black Lives Matter activists created social media hashtags and organized in-person protests (Jackson et al. 2020), aggregators on social media sites promoted the movement, and journalists responded to the increased attention by changing how they wrote about police violence (Freelon, McIlwain, and Clark 2016; Zuckerman et al. 2019). When news media gave more attention to stories about protests and the Black Lives Matter movement, people adapted, too, showing up at protests in greater numbers and posting about it on social media, influencing the algorithms further (De Choudhury et al. 2016). These spirals of attention are so powerful that activists have created special advocacy software to coordinate cycles of feedback during campaigns (Wardle 2014).

Patterns of Impact from Human-Algorithm Feedback

How can we prevent feedback that amplifies injustice while also using it to create power for change toward a just society? Experts have proposed impact assessments as a process for governing algorithms (Reisman et al. 2018). These impact assessments would involve identifying the impacts of feedback, deciding whether those impacts are good or bad, assigning responsibility, and deciding what actions to take (Moss et al. 2021).

Successful assessments start with the ability to identify and name impacts. To identify discrimination by decision makers, researchers created a measurable statistical concept of "bias" (Becker 1957; Narayanan 2018). Although bias is sometimes used to deflect support for accountability (Daumeyer et al. 2019), it has also become a basic tool for regulating discrimination (Barocas, Hardt, and Narayanan 2019).

Can we create new ways to assess the impacts of human-algorithm feedback that are at least as useful and powerful as the idea of bias? Assigning responsibility and providing guidance on change can be difficult when humans and algorithms are continuously changing (Kitchin 2017). But some patterns of impact are now common enough that people have started to name them, even if we don't yet fully understand how they happen.

Reinforcing

When people worry about feedback trapping people in racist, sexist, or extremist views of the world, they are concerned about personalized feedback that reinforces a person's beliefs and behaviors. For example, algorithms may reinforce stereotypes created by primarily-white media industries that already propagate images of people of color as violent and dangerous. Because personalized algorithms make each person's behavior more consistent with their past (Arnold, Chauncey, and Gajos 2020; Negroponte 1996), reinforcing algorithms can entrench epistemic injustice and make personal change more difficult.

Moreover, when teens of color search for media that feature people who look like them, algorithms may show them more results reinforce dominant stereotypes and compound negative self-images (Epps-Darling 2020).

Herding

When Black Lives Matter activists built a movement through local organizing and hashtags, their message was amplified in part by aggregator algorithms (Jackson et al. 2020; Resnick and Varian 1997). These aggregators amplify herding behavior by informing and encouraging people to do something already popular (Broussard 2019; Salganik, Dodds, and Watts 2006). Even before software recommendations, adaptive systems like the Billboard charts encouraged people to flock to popular songs. But herding can be a risk for marginalized groups. When enough people engage in collective discrimination, adaptive systems such as Chicago's policing software can further entrench injustice (Brayne 2020). When discrimination goes viral in the media, the consequences include racial trauma and further marginalization (Bravo et al. 2019). And in societies that normalize violence toward women, aggregator algorithms have further inflamed harassment mobs (Massanari 2017; Yasseri et al. 2016).

Suppressing

Advertising markets sometimes learn to violate employment law by showing fewer job and housing opportunities to women and people of color (Sweeney 2013). If a history of exclusion drives people away from certain opportunities, an algorithm might learn to hide the opportunity altogether. Aggregators can also suppress the interests of a minority group in favor of those of a dominant group. On predominantly white online platforms like Reddit, algorithms have adapted to suppress the views of people of color even in spaces created by and for communities of color (Harmon 2019; Matias, Szalavitz, and Zuckerman 2017). However, algorithmic suppression can also be a tool in service of social justice, such as when it is used to reduce the spread of misinformation designed to dissuade people of color from voting (Funke 2018).

Clustering

When people who have never met each other visit similar websites or act in similar ways online, algorithms can nonetheless treat them as groups. These clusters can be helpful in connecting people with vital advice and opportunities, such as people with common medical conditions who come together to advocate for health equity (Wicks et al. 2010). But clustering can also be dangerous when people have interests that are hateful and violent, and algorithms group them in ways that grow hate groups (Paul 2021; Tufekci 2018).

Dividing

Human-algorithm feedback can further divide society into opposing groups. The power to influence group dynamics has long been a basic tool for political organizing, whether organizers are building power within marginalized groups (Squires 2002) or developing broad coalitions for change (McGhee 2021). With algorithmic polarization, however, groups become more and more socially separated, more opposed

to each other, and less understanding of each other as humans (Finkel et al. 2020). The role of algorithms in social division has been hard to study because polarization and racial resentment in the United States predate algorithms and are maintained by powerful politicians and media corporations that benefit from conflict and hatred (Benkler, Faris, and Roberts 2018; Nyhan 2021).

Optimizing

Pricing algorithms can reinforce economic injustice when they steer people toward behavior that benefits others against their own interests. For instance, algorithms designed by Uber prioritize corporate profits and rider convenience by nudging drivers to take fewer breaks and accept less well-paid work (Scheiber 2017). When these market-management algorithms adapt to the competing preferences of multiple groups at once—for example, customers, workers, and platform owners—they can further entrench discrimination, charging higher fares to people who live in non-white neighborhoods (Pandey and Caliskan 2021). On the positive side, well-designed algorithms could instead optimize for algorithmic reparations, steering prejudiced societies toward economic justice despite themselves (Abebe and Goldner 2018; Davis, Williams, and Yang 2021).

Knowing How to Create Effective Change

Impact assessments also need to include recommendations. To make these suggestions, we need usable knowledge about what kinds of power will lead to meaningful change. That's a problem because scientists can't yet reliably provide that knowledge (Bak-Coleman et al. 2021), although we can describe what it would look like.

In 2012, when Instagram responded to advocacy groups on mental health, self-harm, and eating disorders, they faced a herding problem similar to Reddit's disaster with Jennifer Lawrence's stolen pictures. Instagram tried to make users safer by changing its algorithm to restrict harmful searches. Instead, promoters of self-harm and eating disorders adapted to the changes and gained popularity (Chancellor et al. 2016). Instagram could change the source code of its software, but it couldn't change the code of human culture. In the absence of reliable evidence on effective interventions, activists unknowingly pushed Instagram to make the platform more harmful to young people.

> *If we change algorithms without changing underlying behavior, algorithm policies could fail in similar ways.*

Feedback is hard to change because change might need to come from the algorithm, from humans, or both. It's an algorithm-specific version of a common problem faced by policymakers. In the mid-twentieth century, when civil rights activists ended school segregation through laws and court cases, they hoped to mitigate inequality by changing racist education policies. But policy alone couldn't change the underlying racism of American society. Today, more than sixty years after *Brown v. Board of Education of Topeka*,

most students of color in the United States attend racially segregated schools that receive far fewer resources than primarily white schools (McGhee 2021). If we change algorithms without changing underlying behavior, algorithm policies could fail in similar ways.

Feedback is also hard to change because adaptive algorithms are less predictable than laws or institutions. Since the algorithms respond quickly to changing surroundings, attempts to change feedback patterns can have unpredictable consequences. One group of public health experts recently wrote that "we lack the scientific framework we would need to answer even the most basic questions that technology companies and their regulators face" (Bak-Coleman et al. 2021, 2). They argued that it's currently impossible to tell whether a given algorithm will "promote or hinder the spread of misinformation" (Bak-Coleman et al. 2021, 2). Some scholars even claim that it might be impossible to attain general knowledge about how to change algorithm behavior (Kitchin 2017). Other scientists see this as an advantage, arguing that the consistency of human behavior is the greater barrier to equality. They argue that if humans cannot end patterns of discrimination, it might be better to end injustice by giving more power to algorithms (Mullainathan 2019).

Companies point to these disagreements when trying to avoid regulation (Orben 2020). According to Facebook's Nick Clegg, since predictable algorithms respond to unpredictable humans, the company bears less responsibility for how its algorithms behave. To improve safety, Clegg argues, Facebook should monitor people more closely and enforce policies on humans rather than regulate algorithms (Clegg 2021).

So how can we develop the knowledge needed to advance justice and reduce the harms of human-algorithm feedback? When creating new knowledge, researchers differentiate between general and context-specific knowledge.

Context-specific knowledge can help advocates and policymakers monitor and respond to problems as they happen. Tech companies like Facebook use context-specific knowledge when they monitor the actions of billions of people in real time to decide what content to remove and which accounts to ban (Gillespie 2018; Matias et al. 2020). But companies have strongly resisted attempts to make these algorithms transparent. Even if governments required algorithm inspections similar to car inspections, we do not yet have the testing technology or the staff to evaluate the many adaptive algorithms in use today.

General knowledge can help advocates and governments develop policy solutions that work in more than one situation and for more than one algorithm. Computer simulations of how algorithms will behave in the world, mathematical proofs, and social science theories of behavior might potentially provide this knowledge. This general knowledge could help people prevent injustice and advance a positive vision rather than simply monitor and respond to problems as they happen. For example, general knowledge about the phenomena of self-harm or eating disorders might help guide policymakers to devise effective interventions that prevent the spread of harmful content.

Creating reliable knowledge about feedback should be an urgent priority for everyone who cares about

social justice. Technology makers are introducing algorithms into the world at a massive scale without the ability to predict their social impact. Without new scientific breakthroughs on human-algorithm feedback, we face the risk that interventions for social justice could be ineffective or cause even more harm.

What Can We Do About Human-Algorithm Feedback?

The search for effective interventions is just getting started. Here are some of the most commonly discussed actions that impact assessments could recommend.

The simplest option is to ban the algorithm. Bans are appropriate when the harms strongly outweigh the benefits of a system and when no one has (yet) invented a way to manage those harms effectively. In 2020, for example, Santa Clara, California, became the first US city to ban the use of predictive policing algorithms, citing how they amplify injustice (Asher-Schapiro 2020). But in other contexts, banning algorithms can introduce new harms. For example, because content moderation algorithms can protect marginalized groups from hate online, banning these systems for their known discrimination problems (Davidson, Bhattacharya, and Weber 2019; Dias Oliva, Antonialli, and Gomes 2021) could also expose millions of people to violence and racist threats. Governments sometimes require algorithm operators to remove records on harmful human behavior before algorithms can learn from it. This poorly compensated content moderation work, done by hundreds of thousands of people together with further algorithms, can take a heavy toll on mental health (Roberts 2019).

Photo by Polina Tankilevitch

If an algorithm routinely participates in harmful cycles, designers could change the algorithm (Stray 2021). Companies frequently publish claims that they have adjusted their algorithms in response to public concerns (Bossetta 2020; Hansell 2007). Yet without systematically collected, publicly available evidence on the effects of those changes, it is impossible to know what those changes have accomplished.

Another way to govern feedback is to exclude people from environments where they might influence an algorithm in harmful ways. Between 2019 and 2021, Reddit "quarantined" and then banned several communities with a history of manipulating algorithms to amplify hatred (Isaac 2020; Menegus 2016). One community of color on Reddit excludes white people from some conversations to protect the aggregation algorithm from non-Black voices and votes (Harmon 2019). Predictive decision-making algorithms, when designed and deployed by marginalized communities themselves, often inform these bans and exclusions (Geiger 2016; yellowmix 2015; Zhang et al. 2018).

What if researchers could reliably determine the risk posed by an algorithm before putting it into the world? Potentially harmful algorithms could be tested in the lab before being allowed to enter the market (Ohm and Reid 2016; Tutt 2017). For example, recommendation algorithms can be tested with simulation software that mimics people reading a news feed (Ie et al. 2019; Lucherini et al. 2021; Wainwright and

Eckersley 2021). As with all lab research, the effectiveness of these tests will depend on how realistic they are.

When feedback can't be tested in the lab, one option is to respond to problems after they occur. In April 2021, the Federal Trade Commission announced that it will consider enforcing laws prohibiting unfair or deceptive practices against companies that produce or use "racially biased algorithms" (Jillson 2021). Yet the government may struggle to prove deception if an algorithm's creators themselves can't reliably predict how an algorithm will behave. In the stock market, authorities have introduced "circuit breaker" regulations that monitor markets and temporarily limit or even halt the trading of stocks when signs of volatile feedback emerge (US Securities and Exchange Commission 2020). Similarly, some governments shut down the internet entirely during elections when they expect violence or opposition—policies that human rights advocates widely criticize (Freyburg and Garbe 2018; Howard, Agarwal, and Hussain 2011; Kathuria et al. 2018; West 2016).

How can we learn about the safety of an algorithm or the effectiveness of a policy in real-world situations—while still limiting the risks? Field experiments, when conducted with public consent and careful oversight, enable researchers to audit systems and test policies in controlled circumstances with as few people as possible (Matias, Ko, and Mou 2016; Rey-Mazón et al. 2018). Since the 1970s, researchers and regulators have used audit studies to study discrimination by humans and algorithms (Barocas, Hardt, and Narayanan 2019; Page 2007). Field experiments also provide essential evidence on which policies are effective and which ones backfire. In one study, for example, a community learned effective ways to prevent harmful content from being promoted by Reddit's algorithm—without needing to alter the underlying software (Matias 2020b).

Might restricting data collection make people safer (**Zuboff 2019**)? While data governance is an important policy area, there is no evidence that reducing the information available to algorithms will steer feedback in beneficial ways. Restrictions in data collection can also lead to color-blind policies that can hinder social justice in the same way that bans on affirmative action reduce the toolbox of equity (Bonilla-Silva 2006; Powell 2008).

Advocates sometimes argue that algorithm operators could prevent problems by hiring diverse teams. They argue that the technology industry's history of discrimination (Hicks 2017; Kreiss et al. 2020) has hindered companies' ability to protect marginalized groups (Daniels, Nkonde, and Mir 2019). Yet new employees responsible for ethics and safety can struggle to create changes best led by more senior leaders, especially if new hires are considered outsiders (Dunbar-Hester 2019; Kreiss et al. 2020; Metcalf and Moss 2019; Silbey 2009).

Who Does Governance?

All impact assessments fundamentally include a process of consulting with affected groups to understand the issues and make recommendations (Moss et al. 2021). Why does it matter who gets a voice in the process? Debates about governance always involve struggles over who is responsible and who should be trusted with the power to intervene (Dietz, Ostrom, and Stern 2003; Matias 2015).

The city of Chicago faced one of these struggles over the Strategic Subjects List—a conflict that involved government, companies, universities, and citizen groups. The software was initially commissioned by the city government, developed by the Illinois Institute of Technology, and funded by the US Department of Justice (Asher and Arthur 2017). But affected communities had to force designers and governments to listen to them (Neves 2017). The city halted the program in 2020 after years of lawsuits, research studies documenting its problems, and public pressure (Charles 2020).

Policy debates rarely include the people affected by an issue, although they have the most at stake and the greatest understanding of the context. Yet communities do sometimes have significant governance power, especially when designers give communities tools to adjust and manage algorithms in context (Matias 2019; Matias and Mou 2018). Algorithm designers also sometimes involve communities in the design and training of systems (Halfaker and Geiger 2020). In Chicago, researchers and community organizations coordinated with formerly gang-involved youth to develop an alternative to the city's Strategic Subjects List (Frey et al. 2020). Affected communities have also pioneered algorithm monitoring and accountability, often out of necessity (Matias 2015; Matias and Mou 2018).

Civil society groups also contribute to the governance of algorithmic feedback. Activists and nonprofits organize to influence governance indirectly through lawsuits, research, lobbying campaigns, and many other tactics. By telling people's stories and conducting investigations (Diakopoulos 2014), journalists create scandals that alert policymakers about problems and pressures companies to change (Bossetta 2020).

When algorithm operators like Instagram try to manage human-algorithm feedback by changing code or increasing human surveillance, they are doing governance. Public pressure has also created a market for risk-management businesses that manage other companies' algorithm problems. For example, the UK government is supporting the development of UK-based, for-profit companies to create and rapidly scale safety-focused technologies (Department for Digital Culture, Media, and Sport 2020). Unfortunately, algorithm operators and risk management companies rarely evaluate their policies or publish the results, making it difficult to know if their actions are beneficial (Matias, Ko, and Mou 2016).

Research plays a critical role in governance, alerting people to problems and enabling us to test products and evaluate policies. But not all research advances the public interest. Governments and corporations often announce high-profile hires or donate money to universities to deflect criticism while avoiding making any real changes (Silbey 2009; Weiss 1979). Industry research partnerships can guide wise decisions, but the public is often (rightly) distrustful of industry-funded research (Johnson 2019). Just as in food safety and environmental protection, independent research is essential for understanding and governing human-algorithm feedback (Matias 2020a).

Creating and sustaining social justice will require powerful coalitions of advocates, technology employees, researchers, journalists, and policymakers who combine efforts to understand, reimagine, test, and change complex systems.

Effective governance often involves conflict between different actors (Dietz, Ostrom, and Stern 2003). But social movements can also develop collective power across groups. Whether they use the language of reform, decolonization, or abolition, these movements succeed by organizing many different kinds of activities for justice (Benjamin 2019; Costanza-Chock 2020; Couldry and Mejias 2020). Creating and sustaining social justice will require powerful coalitions of advocates, technology employees, researchers, journalists, and policymakers who combine efforts to understand, reimagine, test, and change complex systems.

The Future of Impact Assessments

When algorithms and humans adapt to each other at scale, the resulting patterns have powerful consequences for every part of our lives. For Wilmer Catalan-Ramirez, who was wrongfully detained, it led to ten months in prison and injuries that could leave him paralyzed for life. For millions of people on Reddit, it spread a culture of sexism and disrespect for women that is already endemic in society. Yet feedback is also a basic building block of movements for social change.

Anyone trying to advance social justice should consider these three basic points about feedback and how to assess its impact:

1. Impact assessments will fail if they focus exclusively on algorithms without considering the underlying human causes of problems.
2. Impact assessments can do more harm than good without new kinds of knowledge: both general science to guide policy-making and monitoring systems to spot problems as they occur.
3. Impact assessments must include affected communities as equal partners in understanding and solving problems.

Acknowledgments

We are grateful to Desmond U. Patton, William Frey, Elizabeth Anne Watkins, Jonathan Stray, Ulises Mejias, Jasmine McNealy, and Daniel Kreiss for early conversations about this article.

Recommended Readings

Abebe, Rediet, and Kira Goldner. 2018. “Mechanism Design for Social Good.” ArXiv:1810.09832 [Cs], October. https://arxiv.org/abs/1810.09832

Kitchin, Rob. 2017. “Thinking Critically about and Researching Algorithms.” Information, Communication & Society 20 (1): 14–29.

Massanari, Adrienne. 2017. “#Gamergate and The Fappening: How Reddit’s Algorithm, Governance, and Culture Support Toxic Technocultures.” New Media & Society 19 (3): 329–46.

Moss, Emanuel, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. 2021. “Assembling Accountability: Algorithmic Impact Assessment for the Public Interest.” Data & Society Research Institute. https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/.

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

References

Abebe, Rediet, and Kira Goldner. 2018. “Mechanism Design for Social Good.” *ArXiv:1810.09832 [Cs]*, October. https://doi.org/10.1145/3284751.3284761.

Alexander, Michelle. 2012. *The New Jim Crow*. New York: The New Press.

Anderson, Carol. 2018. *One Person, No Vote: How Voter Suppression Is Destroying Our Democracy*. New York: Bloomsbury.

Arnold, Kenneth C., Krysta Chauncey, and Krzysztof Z. Gajos. 2020. “Predictive Text Encourages Predictable Writing.” In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 128–38.

Asher, Jeff, and Rob Arthur. 2017. “Inside the Algorithm That Tries to Predict Gun Violence in Chicago.” *The New York Times*, June 13, 2017. https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html.

Asher-Schapiro, Avi. 2020. “California City Bans Predictive Policing in U.S. First.” *Reuters*, June 24, 2020.https://www.reuters.com/article/us-usa-police-tech-trfn-idUSKBN23V2XC.

Bak-Coleman, Joseph B., Mark Alfano, Wolfram Barfuss, Carl T. Bergstrom, Miguel A. Centeno, Iain D. Couzin, Jonathan F. Donges, et al. 2021. “Stewardship of Global Collective Behavior.” *Proceedings of the National Academy of Sciences*, 118 (27): 1–10. https://doi.org/10.1073/pnas.2025764118.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org.

Becker, Gary S. 2010. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Wiley.

Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Bonilla-Silva, Eduardo. 2013. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*. Lanham: Rowman and Littlefield Publishers.

Bossetta, Michael. 2020. "Scandalous Design: How Social Media Platforms' Responses to Scandal Impacts Campaigns and Elections." *Social Media + Society* 6 (2): 1–4. https://doi.org/10.1177/2056305120924777.

Bravo, Diamond Y., Julia Jefferies, Avriel Epps, and Nancy E. Hill. 2019. "When Things Go Viral: Youth's Discrimination Exposure in the World of Social Media." In *Handbook of Children and Prejudice*, 269–87. Cham, Switzerland: Springer.

Brayne, Sarah. 2020. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York: Oxford University Press, USA.

Broussard, Meredith. 2019. *Artificial Unintelligence: How Computers Misunderstand the World*. Illustrated edition. Cambridge, MA: The MIT Press.

Chancellor, Stevie, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. "#Thyghgapp: Instagram Content Moderation and Lexical Variation in pro-Eating Disorder Communities." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1201–13.

Charles, Sam. 2020. "CPD Decommissions 'Strategic Subject List.'" *Chicago Sun-Times*, January 27, 2020.https://chicago.suntimes.com/city-hall/2020/1/27/21084030/chicago-police-strategic-subject-list-party-to-violence-inspector-general-joe-ferguson.

Chavez, Leo R. 2012. *Shadowed Lives: Undocumented Immigrants in American Society*. Belmont, CA: Cengage Learning.

Clegg, Nick. 2021. "You and the Algorithm: It Takes Two to Tango." Medium. March 31, 2021.https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2.

Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: The MIT Press.

Couldry, Nick, and Ulises A. Mejias. 2020. *The Costs of Connection: How Data Are Colonizing Human Life and Appropriating It for Capitalism*. Oxford University Press.

Daniels, Jessie, Mutale Nkonde, and Darakhshan Mir. 2019. "Advancing Racial Literacy in Tech." *Relatório Do Data & Society Fellowship Program*.

Daumeyer, Natalie M., Ivuoma N. Onyeador, Xanni Brown, and Jennifer A. Richeson. 2019. "Consequences of Attributing Discrimination to Implicit vs. Explicit Bias." *Journal of Experimental Social Psychology*, 84 (September): 103812. https://doi.org/10.1016/j.jesp.2019.04.010.

Davidson, Thomas, Debasmita Bhattacharya, and Ingmar Weber. 2019. "Racial Bias in Hate Speech and Abusive Language Detection Datasets." *ArXiv Preprint ArXiv:1905.12516*.

Davis, Jenny L., Apryl Williams, and Michael W. Yang. 2021. "Algorithmic Reparation." *Big Data & Society* 8 (2): 1–12. https://doi.org/10.1177/20539517211044808.

De Choudhury, Munmun, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. "Social Media Participation in an Activist Movement for Racial Equality." In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 10.

Department for Digital Culture, Media, and Sport. 2020. "Online Harms White Paper." United Kingdom: HM Government. https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper.

Diakopoulos, Nicholas. 2015. "Algorithmic Accountability." *Digital Journalism* 3:3: 398-415. https://doi.org/10.1080/21670811.2014.976411

Dias Oliva, Thiago, Dennys Marcelo Antonialli, and Alessandra Gomes. 2021. "Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online." *Sexuality & Culture* 25 (2): 700–732. https://doi.org/10.1007/s12119-020-09790-w.

Dietz, Thomas, Elinor Ostrom, and Paul C. Stern. 2003. "The Struggle to Govern the Commons." *Science* 302 (5652): 1907–12.

Donohoe, Martin T. 2012. *Public Health and Social Justice: A Jossey-Bass Reader*. John Wiley & Sons.

Dunbar-Hester, Christina. 2019. *Hacking Diversity: The Politics of Inclusion in Open Technology Cultures*. Vol. 21. Princeton, NJ: Princeton University Press.

Ekstrand, Michael D., John T. Riedl, and Joseph A. Konstan. 2011. "Collaborative Filtering Recommender Systems." *Foundations and Trends® in Human–Computer Interaction* 4 (2): 81–173.https://doi.org/10.1561/1100000009.

Epps-Darling, Avriel. 2020. “Racist Algorithms Are Especially Dangerous for Teens.” *The Atlantic*, October 24, 2020. https://www.theatlantic.com/family/archive/2020/10/algorithmic-bias-especially-dangerous-teens/616793/

Epps-Darling, Avriel, Romain Takeo Bouyer, and Henriette Cramer. 2020. “Artist Gender Representation in Music Streaming.” In *Proceedings of the 21st International Society for Music Information Retrieval Conference (Montréal, Canada)(ISMIR 2020). ISMIR*, 248–54.

Finkel, Eli J., Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary C. McGrath, Brendan Nyhan, and David G. Rand. 2020. “Political Sectarianism in America.” *Science* 370 (6516): 533–36.

Freelon, Deen, Charlton D. McIlwain, and Meredith Clark. 2016. “Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice.” *Center for Media & Social Impact, American University, Forthcoming*. https://cmsimpact.org/resource/beyond-hashtags-ferguson-blacklivesmatter-online-struggle-offline-justice/.

Frey, William R., Desmond U. Patton, Michael B. Gaskell, and Kyle A. McGregor. 2020. “Artificial Intelligence and Inclusion: Formerly Gang-Involved Youth as Domain Experts for Analyzing Unstructured Twitter Data.” *Social Science Computer Review* 38 (1): 42–56.

Freyburg, Tina, and Lisa Garbe. 2018. “Blocking the Bottleneck: Internet Shutdowns and Ownership at Election Times in Sub-Saharan Africa.” *International Journal of Communication* 12: 3896–3916.

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

Funke, Daniel. 2018. “Facebook Is Now Downranking Stories with False Headlines.” *Poynter*, October 24, 2018.https://www.poynter.org/fact-checking/2018/facebook-is-now-downranking-stories-with-false-headlines/.

Geiger, R. Stuart. 2016. “Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space.” *Information, Communication & Society* 19 (6): 787–803.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Greenberg, Andy. 2014. “Hacked Celeb Pics Made Reddit Enough Cash to Run Its Servers for a Month.” *Wired*, September 10, 2014. https://www.wired.com/2014/09/celeb-pics-reddit-gold/.

Halfaker, Aaron, and R. Stuart Geiger. 2020. “Ores: Lowering Barriers with Participatory Machine

Learning in Wikipedia." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2): 1–37.

Hansell, Saul. 2007. "Google Keeps Tweaking Its Search Engine." *The New York Times*, June 3, 2007.https://www.nytimes.com/2007/06/03/business/yourmoney/03google.html.

Harmon, Amy. 2019. "Discussing Blackness on Reddit? Photograph Your Forearm First." *The New York Times*, October 8, 2019. https://www.nytimes.com/2019/10/08/us/reddit-race-black-people-twitter.html.

Hicks, Mar. 2017. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: MIT Press.

Howard, Philip N., Sheetal D. Agarwal, and Muzammil M. Hussain. 2011. "The Dictators' Digital Dilemma: When Do States Disconnect Their Digital Networks?" Issues in Technology Innovation. Brookings Institution.https://www.brookings.edu/wp-content/uploads/2016/06/10_dictators_digital_network.pdf.

Ie, Eugene, Chih-wei Hsu, Martin Mladenov, Vihan Jain, Sanmit Narvekar, Jing Wang, Rui Wu, and Craig Boutilier. 2019. "RecSim: A Configurable Simulation Platform for Recommender Systems." *ArXiv:1909.04847 [Cs, Stat]*, September. http://arxiv.org/abs/1909.04847.

"Investigation of the Chicago Police Department." 2017. United States Department of Justice Civil Rights Division and United States Attorney's Office Northern District of Illinois.https://www.justice.gov/opa/file/925846/download.

Isaac, Mike. 2020. "Reddit, Acting Against Hate Speech, Bans 'The_Donald' Subreddit." *The New York Times*, June 29, 2020. https://www.nytimes.com/2020/06/29/technology/reddit-hate-speech.html.

Jackson, Sarah J., Moya Bailey, Brooke Foucault Welles, and Genie Lauren. 2020. *#HashtagActivism: Networks of Race and Gender Justice*. Illustrated edition. Cambridge: The MIT Press.

Jillson, Elisa. 2021. "Aiming for Truth, Fairness, and Equity in Your Company's Use of AI." *Federal Trade Commission* (blog). April 19, 2021. https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

Johnson, Courtney. n.d. "Most Americans Are Wary of Industry-Funded Research." Pew Research Center, October 4, 2019. https://www.pewresearch.org/fact-tank/2019/10/04/most-americans-are-wary-of-industry-funded-research/.

Kashner, Sam. 2014. "Exclusive: Jennifer Lawrence Speaks About Her Stolen Photos." *Vanity Fair*, October 20, 2014. https://www.vanityfair.com/hollywood/2014/10/jennifer-lawrence-photo-hacking-privacy.

Kathuria, Rajat. 2018. *The Anatomy of an Internet Blackout: Measuring the Economic Impact of Internet Shutdowns in India*. Indian Council for Research on International Economic Relations.

Kitchin, Rob. 2017. "Thinking Critically about and Researching Algorithms." *Information, Communication & Society* 20 (1): 14–29.

Ko, Allan, Merry Mou, and Nathan Matias. 2016. "The Obligation To Experiment." *MIT Media Lab* (blog). 2016.https://medium.com/mit-media-lab/the-obligation-to-experiment-83092256c3e9.

Kreiss, Daniel, Kirsten Adams, Jenni Ciesielski, Haley Fernandez, Kate Frauenfelder, and Brinley Lowe. 2020. *Recoding the Boys' Club: The Experiences and Future of Women in Political Technology*. Oxford University Press.

Li, Lihong, Wei Chu, John Langford, and Robert E. Schapire. 2010. "A Contextual-Bandit Approach to Personalized News Article Recommendation." In *Proceedings of the 19th International Conference on World Wide Web*, 661–70. WWW '10. New York, NY, USA: Association for Computing Machinery.https://doi.org/10.1145/1772690.1772758.

Lucherini, Eli, Matthew Sun, Amy Winecoff, and Arvind Narayanan. 2021. "T-RECS: A Simulation Tool to Study the Societal Impact of Recommender Systems." *ArXiv:2107.08959 [Cs]*, July. http://arxiv.org/abs/2107.08959.

Massanari, Adrienne. 2017. "# Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures." *New Media & Society* 19 (3): 329–46.

Matias, J. Nathan. 2015. "The Tragedy of the Digital Commons." *The Atlantic*, June 8, 2015.https://www.theatlantic.com/technology/archive/2015/06/the-tragedy-of-the-digital-commons/395129/.

Matias, J. Nathan. 2017. "Governing Human and Machine Behavior in an Experimenting Society." PhD Thesis, Massachusetts Institute of Technology.

Matias, J. Nathan. 2019. "The Civic Labor of Volunteer Moderators Online." *Social Media + Society* 5 (2): 2056305119836778. https://doi.org/10.1177/2056305119836778.

Matias, J. Nathan. 2020a. "Why We Need Industry-Independent Research on Tech & Society." *Citizens and Technology Lab* (blog). January 7, 2020. https://citizensandtech.org/2020/01/industry-independent-research/.

Matias, J. Nathan. 2020b. "Nudging Algorithms by Influencing Human Behavior: Effects of Encouraging Fact-Checking on News Rankings," March. https://doi.org/DOI 10.17605/OSF.IO/M98B6.

Matias, J. Nathan, and Merry Mou. 2018. "CivilServant: Community-Led Experiments in Platform

Governance.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 9. ACM.

Matias, J. Nathan, Merry Ember Mou, Jonathon Penney, and Maximilian Klein. 2020. “Do Automated Legal Threats Reduce Freedom of Expression Online? Preliminary Results from a Natural Experiment,” September.https://doi.org/DOI 10.17605/OSF.IO/NC7E2.

Matias, J. Nathan, Sarah Szalavitz, and Ethan Zuckerman. 2017. “FollowBias: Supporting Behavior Change toward Gender Equality by Networked Gatekeepers on Social Media.” In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1082–95.

McGhee, Heather. 2021. *The Sum of Us: What Racism Costs Everyone and How We Can Prosper Together*. New York: One World.

Menegus, Bryan. 2016. “Reddit Is Tearing Itself Apart.” *Gizmodo*, January 29, 2016. http://gizmodo.com/reddit-is-tearing-itself-apart-1789406294.

Metcalf, Jacob, and Emanuel Moss. 2019. “Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics.” *Social Research: An International Quarterly* 86 (2): 449–76.

Moreno, Nereida. 2017. “Chicago Settles Suit with Immigrant Falsely Accused of Gang Ties – Chicago Tribune.” *Chicago Tribune*, December 7, 2017. https://www.chicagotribune.com/news/ct-met-immigration-lawsuit-settled-1206-story.html.

Moss, Emanuel, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. 2021. “Assembling Accountability: Algorithmic Impact Assessment for the Public Interest.” Data & Society Research Institute. https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/.

Mullainathan, Sendhil. 2019. “Biased Algorithms Are Easier to Fix Than Biased People – The New York Times.” *New York Times*, December 6, 2019. https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

Narayanan, Arvind. n.d. “Translation Tutorial: 21 Fairness Definitions and Their Politics” 1170.

Negroponte, Nicholas. 1996. *Being Digital*. 1st edition. New York, NY: Vintage.

Neves, Jessey. 2017. “Man Whose False Inclusion in CPD Gang Database Made Him ICE Target Reaches Settlement with City of Chicago.” *National Immigration Project*, December 6, 2017.https://nipnlg.org/pr/2017_06Dec_wilmer-v-chicago-settlement.html.

Nyhan, Brendan. 2021. “Why the Backfire Effect Does Not Explain the Durability of Political

Misperceptions." *Proceedings of the National Academy of Sciences* 118 (15). https://doi.org/10.1073/pnas.1912440117.

Ohm, Paul, and Blake E. Reid. 2016. "Regulating Software When Everything Has Software." *George Washington Univrsity Law Review*, November. https://papers.ssrn.com/abstract=2873751.

Orben, Amy. 2020. "The Sisyphean Cycle of Technology Panics." *Perspectives on Psychological Science* 15 (5): 1143–57.

Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *The Annals of the American Academy of Political and Social Science* 609 (1): 104–33.

Pandey, Akshat, and Aylin Caliskan. 2021. "Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms." In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 822–33. AIES '21. New York, NY, USA: Association for Computing Machinery.https://doi.org/10.1145/3461702.3462561.

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

Powell, John A. 2008. "Post-Racialism or Targeted Universalism." *Denver Law Review* 86: 785.

Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability." AI Now Institute.

Resnick, Paul, and Hal R. Varian. 1997. "Recommender Systems." *Communications of the ACM* 40 (3): 56–58.

Reverby, Susan M. 2012. *Tuskegee's Truths: Rethinking the Tuskegee Syphilis Study*. Chapel Hill: University of North Carolina Press.

Rey-Mazón, Pablo, Hagit Keysar, Shannon Dosemagen, Catherine D'Ignazio, and Don Blair. 2018. "Public Lab: Community-Based Approaches to Urban and Environmental Health and Justice." *Science and Engineering Ethics* 24 (3): 971–97.

Roberts, Sarah T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven, CT: Yale University Press.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311 (5762): 854–56.

Saunders, Jessica, Priscillia Hunt, and John S. Hollywood. 2016. “Predictions Put into Practice: A Quasi-Experimental Evaluation of Chicago’s Predictive Policing Pilot.” *Journal of Experimental Criminology* 12 (3): 347–71.

Scheiber, Noam. (2017, April 2). “How Uber Uses Psychological Tricks to Push Its Drivers’ Buttons.” *The New York Times*, April 2, 2017. https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html.

Silbey, Susan S. 2009. “Taming Prometheus: Talk about Safety and Culture.” *Annual Review of Sociology* 35: 341–69.

Sparkes, Matthew. 2014. “Nude Celebrity Selfies: The Internet Never Forgets.” *The Telegraph*, September 2, 2014. https://www.telegraph.co.uk/technology/internet-security/11069788/Nude-celebrity-selfies-the-internet-never-forgets.html.

Squires, Catherine R. 2002. “Rethinking the Black Public Sphere: An Alternative Vocabulary for Multiple Public Spheres.” *Communication Theory* 12 (4): 446–68.

Stray, Jonathan. 2021. “Designing Recommender Systems to Depolarize.” ArXiv:2107.04953 [cs.IR]. https://arxiv.org/abs/2107.04953.

Sweeney, Latanya. 2013. “Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising.” *Queue* 11 (3): 10–29.

Tufekci, Zeynep. 2018. “Opinion | YouTube, the Great Radicalizer.” *The New York Times*, March 10, 2018.https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

Tutt, Andrew. 2017. “An FDA for Algorithms.” *Administrative Law Review* 69 (1): 83–124.

U.S. Securities and Exchange Commission. 2020. “Staff Report on Algorithmic Trading in U.S. Capital Markets.”https://www.sec.gov/tm/reports-and-publications/special-studies/algo_trading_report_2020.

Virgilio, Gianluca Piero Maria. 2019. “High-Frequency Trading: A Literature Review.” *Financial Markets and Portfolio Management* 33 (2): 183–208. https://doi.org/10.1007/s11408-019-00331-6.

Wainwright, Carroll L., and Peter Eckersley. 2021. “SafeLife 1.0: Exploring Side Effects in Complex Environments.” *ArXiv:1912.01217 [Cs]*, February. http://arxiv.org/abs/1912.01217.

Wardle, Hilary. 2014. “How New ‘Crowdspeaking’ Site Thunderclap Is Revolutionising Online Awareness Raising.” *HuffPost UK*, March 25, 2014. https://www.huffingtonpost.co.uk/hilary-wardle/how-new-crowdspeaking-sit_b_5021877.html.

Weiss, Carol H. 1979. “The Many Meanings of Research Utilization.” *Public Administration Review* 39 (5): 426–31.

West, Darrell M. 2016. “Internet Shutdowns Cost Countries $2.4 Billion Last Year.” *Center for Technological Innovation at Brookings, Washington, DC*.

Wicks, Paul, Michael Massagli, Jeana Frost, Catherine Brownstein, Sally Okun, Timothy Vaughan, Richard Bradley, and James Heywood. 2010. “Sharing Health Data for Better Outcomes on PatientsLikeMe.” *Journal of Medical Internet Research* 12 (2): e19.

Yasseri, Taha, Helen Margetts, Peter John, and Scott Hale. 2016. *Political Turbulence: How Social Media Shape Collective Action*. Princeton University Press.

yellowmix. 2015. *Introduction – Saferbot*. r/Saferbot. https://www.reddit.com/r/Saferbot/wiki/introduction.

Zhang, Justine, Jonathan Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Dario Taraborelli, and Nithum Thain. 2018. “Conversations Gone Awry: Detecting Early Signs of Conversational Failure.” In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1350–61. Melbourne, Australia: Association for Computational Linguistics.https://doi.org/10.18653/v1/P18-1125.

Zuboff, Shoshana. 2019. “Surveillance Capitalism and the Challenge of Collective Action.” *New Labor Forum* 28, no. 1 (January): 10–29. https://doi.org/10.1177/1095796018819461.

Zuckerman, Ethan, J. Nathan Matias, Rahul Bhargava, Fernando Bermejo, and Allan Ko. 2019. “Whose Death Matters? A Quantitative Analysis of Media Attention to Deaths of Black Americans in Police Confrontations, 2013–2016.” *International Journal of Communication,* 13 (2019): 27.