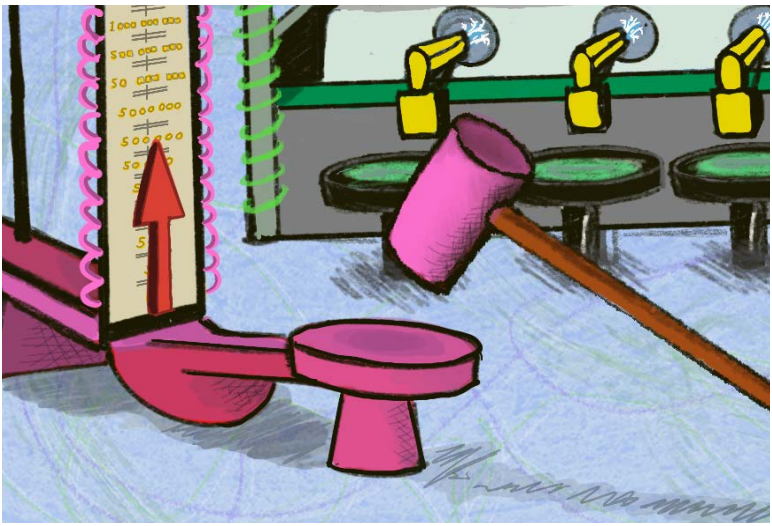
OPTIMIZING FOR WHAT?
ALGORITHMIC AMPLIFICATION AND SOCIETY



Algorithmic Displacement of Social Trust

By Benjamin Laufer and Helen Nissenbaum

KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY

In April 2023, the Knight Institute hosted a symposium, "Optimizing for What? Algorithmic Amplification and Society," to explore how law and policy regulates or should regulate the algorithms that power social media platforms' recommendation systems and the challenges and opportunities that arise from these systems. The symposium was organized in partnership with the Institute's 2022-2023 Visiting Senior Research Scientist Arvind Narayanan and took place at Columbia University.

The essays in this series were originally presented and discussed at this event. Written by leading scholars, nonprofit leaders, technologists, and professionals from a wide range of disciplines, including computer science, psychology, law, social science, philosophy and other fields, these essays offer more precise explanations of how algorithms are designed, deployed, and evaluated and propose interventions that would mitigate some of the harms caused by amplification. This series also engaged with normative questions about algorithmic recommenders: What should they optimize for? How do we design systems to promote healthier public discourse? Who makes these decisions?

The symposium was conceptualized by Knight Institute staff, including Jameel Jaffer, executive director; Katy Glenn Bass, research director; Arvind Narayanan, visiting senior research scientist; Alex Abdo, litigation director; and Larry Siems, chief of staff. The essay series was edited by Glenn Bass and Narayanan with additional support from Lorraine Kenny, communications director; Victoria Tang, legal research fellow; Mia Speier, research coordinator; Kushal Dev, research fellow; Avian Muñoz, intern; and Stephen Dai, intern.

The full series is available at knightcolumbia.org/research/

NEW INFORMATION TECHNOLOGIES are sometimes described as having a "democratizing" effect. Anyone's single-sentence post or 10-second video can potentially reach millions over the internet, meaning that technology has broken barriers that long constrained the reach of individual ideas, opinions, or observations. Yet, paradoxically, there is growing anxiety that our system of communication is posing severe, unprecedented, and even fatal threats to democracy.[1]

In the mid-1990s and early 2000s, new media and internet technologies were celebrated for their potential to advance knowledge, unyoke us from long-standing constraints,[2] and democratize access to communication and speech.[3] A dominant theme was the power of these new media

---

1 For a discussion on the impact of social media on democracy, *see* Sunstein, Cass. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, 2017. For broader discussions on threats to democracy, *see* Bermeo, Nancy. "On Democratic Backsliding." *J. Democracy* 27.1 (2016): 5. *See also* Levitsky, Steven, and Daniel Ziblatt. *How Democracies Die*. Crown, 2018.

2 *See, e.g.*, Johnson, David R., and David Post. "Law and Borders: The Rise of Law in Cyberspace." *Stanford Law Review* 48.5 (1996): 1367–1402.

3 For a description of—and research on–this optimistic view, see Loader, Brian D., and Dan Mercea. "Networking Democracy? Social Media Innovations and Participatory Politics." *Information, Communication & Society* 14.6 (2011): 757–769.

to break the chokehold that old media elites had over what was printed and broadcast, determining what people saw and heard. These gatekeepers decided what stories, experiences, opinions, and artworks were worthy of attention—and what were not. The initial wave of optimism embraced the idea that the new digital medium had the capacity to extend the power to participate in the creation, publication, and distribution of material to an ever-broader spectrum of readers, listeners, and choosers of material.

The passing years have borne witness to an evolving media landscape that is more complicated than these visions, one with a darker edge. Although it is true that almost anyone can publish on the internet, there is no guarantee that anyone will read or hear them. The sheer volume of content online means that gatekeeping is no less important now than it was in pre-internet days.[4] This time, however, the concern is less about who and what gets published and more about who and what gets attention. Belying the ideals of the early optimistic vision of a diversified mediascape, we seem to have reverted to one that is concentrated,[5] vesting the powers of selection, recommendation, and distribution in the hands of a few. But the conditions now are markedly different. Instead of being up to human decision-makers (acting in capacities of chief editors, heads of broadcast media, heads of marketing companies, etc.), creation, selection, and distribution are directed through a patchwork of automated systems and human decisions functioning atop social media platforms, which, for the most part, are owned by a dominant few companies. We refer to these systems and decisions as *algorithmic* when they incorporate formalized, computational procedures, rules, or instructions, which execute some function or purpose.[6]

Framed in a positive light, the use of algorithms to curate, direct, and aggregate content is a way of coping with its massive scale while remaining true to the spirit of democratization. Automating some of these operations

---

4 *See* Shoemaker, Pamela J., and Timothy Vos. *Gatekeeping Theory*. Routledge, 2009.

5 Van Couvering, Elizabeth. "New Media? The Political Economy of Internet Search Engines." *Annual Conference of the International Association of Media & Communications Researchers* (2004).

6 *See* Gillespie, Tarleton. "The Relevance of Algorithms." *Media Technologies: Essays on Communication, Materiality, and Society* (2014): 167.

through algorithmic systems can enable participation and representation in new and exciting ways, even though bottlenecks and gatekeeping may still be required. Finding quality shows, movies, books, and restaurants, for example, no longer depends on the published reviews of one or a handful of journalists; instead, masses of people can contribute comments and opinions. Social movements have been empowered by algorithmic systems that connect like-minded individuals—sometimes geographically dispersed—to mobilize against long-standing injustices and abuses of centralized powers.[7]

The downsides of algorithmic curation, however, are becoming increasingly evident, too. Savvy users have been able to exploit algorithmic media systems, for example, to embolden vocal support for far-right, racist ideologies[8] and conspiracy theories.[9] Coordinated attacks on algorithmic systems—including by nation-states—have undermined the reliability of information distributed by historically trustworthy networks, sowing dissent and distrust.[10] Biases embedded in algorithmic systems, sometimes too subtle to easily detect, systematically and deterministically exclude and marginalize victims in new ways, even proliferating novel forms of bigotry.[11] Increasingly polarized societies are vulnerable to fragmentation, as insular subgroups take form with differing values and ideological commitments.[12] Horrific acts

---

7 *See* Mendes, Kaitlynn, Jessica Ringrose, and Jessalynn Keller. "#MeToo and the Promise and Pitfalls of Challenging Rape Culture through Digital Feminist Activism." *European Journal of Women's Studies* 25.2 (2018): 236–246. *See also* Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, 2017.

8 *See, e.g.*, Jakubowicz, Andrew. "Alt_Right White Lite: Trolling, Hate Speech and Cyber Racism on Social Media." *Cosmopolitan Civil Societies: An Interdisciplinary Journal* 9.3 (2017): 41–60. For empirical work on the spread of hate speech on the social media platform Gab, *see* Mathew, Binny, et al. "Spread of Hate Speech in Online Social Media." *Proceedings of the 10th ACM Conference on Web Science* (2019).

9 De Zeeuw, Daniel, et al. "Tracing Normiefication: A Cross-Platform Analysis of the QAnon Conspiracy Theory." *First Monday* 25.11 (2020).

10 Tucker, Joshua A., et al. "From Liberation to Turmoil: Social Media and Democracy." *J. Democracy* 28 (2017): 46.

11 Noble, Safiya Umoja. *Algorithms of Oppression*. New York University Press, 2018. Benjamin, Ruha. *Race after Technology*. Polity Press, 2019. *See also* Phan, Thao, and Scott Wark. "Racial Formations as Data Formations." *Big Data & Society* 8.2 (2021).

12 For a much-discussed example, see Jamieson, Kathleen Hall, and Joseph N. Cappella. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press, 2010.

of violence have been carried out, inspired by algorithmically steered vicious, divisive, and skewed content.[13]

Motivated by mounting concerns over the effects of algorithmically circulating online content on dangerous behaviors and corrosive social formations, scholars and journalists have reached for the concept of "algorithmic amplification" as both an explanation and a warning. Experts in the areas of media and technology have long posed questions about "amplification," about how and why certain content, information, viewpoints, or even simply, names spread both broadly and unevenly.[14] In the face of the further proliferation of hate speech, mis- and disinformation, and the acute rise in bias and polarization,[15] experts are taking novel lines of investigation into the ascent of algorithmic systems as dominant engines for selecting, distributing, and recommending content.[16]

Our paper takes up this line of questioning, with a focus on what we are calling *problematic* algorithmic amplification. As search, recommendation, and other information and media services are formalized and operationalized algorithmically, the threats they pose do not stop at the proliferation of

---

13 Stray, Jonathan, Ravi Iyer, and Helena Puig Larrauri. "The Algorithmic Management of Polarization and Violence on Social Media." *Knight First Amendment Institute*, August 22, 2023, https://knightcolumbia.org/content/the-algorithmic-management-of-polarization-and-violence-on-social-media.

14 *See* Phillips, Whitney. "The Oxygen of Amplification." *Data & Society*, May 22, 2018, https://datasociety.net/library/oxygen-of-amplification/. For another example of how political views and movements impact policy, using the terminology of "amplification," *see* Agnone, Jon. "Amplifying Public Opinion: The Policy Impact of the US Environmental Movement." *Social Forces* 85.4 (2007): 1593–1620.

15 For evidence of rising affective polarization over time, *see* Iyengar, Shanto, et al. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22 (2019): 129–146. For comparisons across countries, *see* Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. "Cross-Country Trends in Affective Polarization." *Review of Economics and Statistics* (2022): 1–60.

16 Donovan and boyd (2021), for example, use the notion of 'strategic silence' and 'strategic amplification' to discuss editorial and content moderation approaches, in particular, those that are strategically necessary to avoid harmful outcomes. *See* Donovan, Joan, and danah boyd. "Stop the Presses? Moving from Strategic Silence to Strategic Amplification in a Networked Media Ecosystem." *American Behavioral Scientist* 65.2 (2021): 333–350. Riemer and Peter (2021) use the term "algorithmic audiencing" to describe the phenomenon where algorithms determine the audiences reached by speech. *See* Riemer, Kai, and Sandra Peter. "Algorithmic Audiencing: Why We Need to Rethink Free Speech on Social Media." *Journal of Information Technology* 36.4 (2021): 409–426. For a broader discussion of how technology companies perform the functions of media companies while evading qualification as such, *see* Napoli, Philip, and Robyn Caplan. "Why Media Companies Insist They're Not Media Companies, Why They're Wrong, and Why It Matters." *First Monday* (2017).

problematic content, including hate speech, polarization, mis- and disinformation, etc. Though content of this sort is worrying, we seek to identify underlying reasons why algorithmic systems have increased our collective social vulnerability to their proliferation. The pages that follow will lay out our argument, that algorithmic systems, including those found on dominant social media platforms, pose existential threats to entrenched *processes* that perform many of the same functions, namely, selective circulation of information and content.

These processes, which may have evolved over long stretches of time, are *sound* to the extent that they serve societal and institutional purposes. Sound processes, as we have defined them, provide good reason for trusting what we uncover at certain times and for certain needs, whatever the nature of what it is we seek at those times and for those needs. To illustrate these ideas, we will discuss epistemic processes, aimed at achieving trustworthy knowledge, and democratic processes, aimed at the collective determination of governance. Before returning to our discussion of processes and threats from problematic algorithmic amplification, we provide a few definitions and clarifications.

ALGORITHMIC AMPLIFICATION: WHAT IS IT?

OUR RESEARCH SUGGESTS that the term *algorithmic amplification* was not formally introduced but, instead, emerged organically in public discussion. Although the term has been a useful organizing concept for journalists and academics, many who use it agree it is imprecise. For example, Keller (2021), a legal scholar, admits that "the concept of internet amplification may be inevitably fuzzy at the edges."[17] Writing on journalistic practices for reporting on extremists and manipulators, Phillips (2018), a journalism and communications scholar, noted "how fraught questions of amplification really are; just how damned if we do, damned if

---

17 Keller, Daphne. "Amplification and Its Discontents: Why Regulating the Reach of Online Content is Hard." *J. Free Speech* L. 1 (2021): 227.

we don't the landscape can be."[18] Eckles (2021), approaching the topic from statistics and social science, writes:

> Assessing what a ranking algorithm is amplifying is not a trivial task that platforms have simply neglected. And we may be substantially mislead [sic] by assessments of algorithmic amplification that simply compare two rankings of the same content.[19]

The imprecision surrounding *algorithmic amplification* should not be too surprising, considering its origins as a metaphor, drawn from signal processing and dynamical systems, predating the internet and other computational technologies by decades. And before the 20th century, its prosaic meaning was "to make ample,"[20] an expansion or growth of something. With the advent of electrical signals and acoustics, an "amplifier" was a tool that could increase the amplitude of a signal, with a range of applications from live music to electrical circuits.

In the 1950s, amplification entered conversations around cybernetics and systems science. Applied to early efforts in artificial intelligence (AI), the term "intelligence amplification" (IA), coined by Ashby (1956), referred to the potential for information technologies to augment human intelligence and develop knowledge at a faster rate.[21] The idea of IA is contrasted with AI because IA focuses more on augmenting and advancing human knowledge while AI has historically been about replicating intelligence in a machine. This usage suggests societal-scale, complex systems thinking that connects with algorithmic amplification concerns raised today.[22]

---

18 Phillips (2018). For a discussion on manipulation and the specific vulnerabilities of metric-based algorithmic recommendations and disinformation, see Marwick, Alice E., and Rebecca Lewis. "Media Manipulation and Disinformation Online." *Data & Society*, May 15, 2017, https://datasociety.net/library/media-manipulation-and-disinfo-online/.

19 Eckles, Dean. "Algorithmic Transparency and Assessing Effects of Algorithmic Ranking." *Testimony Before the Senate Subcommittee on Communications, Media, and Broadband*, December 9, 2021, https://www.commerce.senate.gov/services/files/62102355-DC26-4909-BF90-8FB068145F18.

20 *See*, e.g., Thomas Kerchever Arnold. Spelling Turned Etymology. Gilbert & Rivington Printers, 1844: 7. *See also* etymonline, "amplify (v.)," https://www.etymonline.com/word/amplify.

21 Ashby, William R. *An Introduction to Cybernetics*. Chapman & Hall, 1956.

22 For a discussion of social media platforms as "complex systems," *see* Narayanan, Arvind. "Un

The term's recent use in describing the distribution of content on social media[23] marks an expansion in meaning. In these contexts, the term is used not to highlight technology's ability to expand knowledge but to warn about its ability to expand the spread of all sorts of undesirable information and activities. Misinformation, extremism, hate speech, bias, and other corrosive social effects have been bolstered by algorithmic systems, tuned to provide relevant or entertaining content. Critics have sourced these corrosive effects to the internet and other digital systems as a way of expressing and making sense of their worries. However, articulations of the problems surrounding amplification suffer from the term's lingering ambiguities. On one hand, an overly broad definition runs the risk of lacking "teeth." If any reproduction of speech is amplification, it becomes difficult to pinpoint why amplification matters. On the other hand, overly narrow definitions, such as "departure from a baseline feed,"[24] may squeeze out amplifying practices that people worry about, by excusing baseline behaviors and choices as neutral.

Is amplification necessarily problematic? At least part of the term's imprecision arises because those who use it have not converged on an answer. Often, "algorithmic amplification" is used both to describe and appraise. As a description, it characterizes and circumscribes various practices without judging them in ethical terms. Normative uses, by contrast, identify these practices as ones deserving of moral scrutiny and valuation, whether

---

derstanding Social Media Recommendation Algorithms." *Knight First Amendment Institute*, March 9, 2023, https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.

23 *See, e.g.*, McCabe, David. "Lawmakers Target Big Tech 'Amplification.' What Does That Mean?" *New York Times*, December 1, 2021, https://www.nytimes.com/2021/12/01/technology/big-tech-amplification.html. Beckett, Louis, and Julia Carrie Wong, "The Misinformation Media Machine Amplifying Trump's Election Lies." *The Guardian*, November 10, 2020, https://www.theguardian.com/us-news/2020/nov/10/donald-trump-us-election-misinformation-media. Whittaker, Joe, Seán Looney, Alastair Reed, and Fabio Votta. "Recommender Systems and the Amplification of Extremist Content." *Internet Policy Review* 10.2 (2021).

24 Measurement of amplification as a departure from a baseline feed lends itself to empirical analysis comparing the relative effects of two different algorithmic ranking logics for online platforms. A common example is Twitter audit studies—*see, e.g.*, Huszár, Ferenc, et al. "Algorithmic Amplification of Politics on Twitter." *Proceedings of the National Academy of Sciences* 119.1 (2022). Milli, Smitha, et al. "Twitter's Algorithm: Amplifying Anger, Animosity, and Affective Polarization." *arXiv preprint arXiv:2305.16941* (2023). Theoretical work has been proposed as well: Cen, Sarah H., Aleksander Madry, and Devavrat Shah. "A User-Driven Framework for Regulating and Auditing Social Media." *arXiv preprint arXiv:2304.10525* (2023).

positive or negative.[25] To achieve greater clarity, this paper teases apart these two uses, first elaborating a descriptive concept of algorithmic amplification, and then providing a normative account, further distinguishing the focus of greatest interest to us, here, as cases of *problematic* algorithmic amplification.

We use the term ***amplification*** to cover the systematic procedures that platforms, publishers, and services use in order to expand the reach of content, either in absolute terms (i.e., across-the-board) or selectively for particular audiences and categories of content. It is worth noting that by virtue of some content being boosted, in relative terms, the neglected content might be considered *suppressed*. Amplification may be further intensified due to the dynamic effects of the initiating steps; for example, people whose beliefs, incentives, and behaviors have been affected by amplified content may further amplify this content in their interactions with one another or with the platform.

We use the term ***algorithmic amplification*** to refer to the amplification of content as a result of formalized and operationalized sets of instructions, typically carried out by computer systems. Although algorithmic amplification involves machine automation, the degree of automation can vary across applications, in some instances requiring significant human involvement and oversight.[26] When the logic dictating the reach of content is expressed as an algorithm, whether or not the content gets amplified depends on component procedures and rules. For example, a social media post that uses a certain tag, or contains text limited to a certain length, may be rewarded through the encoded rules and instructions that are collectively referred to as an algorithm. The advantages of algorithmic amplification to platforms and services that utilize it (and—they might argue—to their users), as with other automated systems, are efficiencies of speed and scale. Algorithmic

---

25 The potentially many positive and negative implications of algorithmic ranking is discussed by Eckles (2021), who proposes certain empirical approaches to begin to tease out these effects.

26 This is sometimes described as "human-in-the-loop." *See, e.g.*, Cranor, Lorrie F. "A Framework for Reasoning about the Human in the Loop." *Proceedings of the 1st Conference on Usability, Psychology, and Security* (2008). For a reframing of this conception that aims to highlight that algorithms are often used to inform systems of human decisions, *see* Green, Ben, and Yiling Chen. "The Principles and Limits of Algorithm-in-the-Loop Decision Making." *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW1 Article 50 (2019): 1–24. In either framing, an algorithmic system is one in which both humans and algorithms play a role in a system.

amplification, implicitly and explicitly, is indubitably baked into systems that are integral to contemporary, digital societies, such as automated web-searching, all manner of matching (e.g., dating), and a great variety of recommendation platforms (e.g., Yelp, Reddit).

PROBLEMATIC ALGORITHMIC AMPLIFICATION

ARMED WITH THIS GENERAL account of algorithmic amplification, we turn to the normative question at the heart of this article: When is algorithmic amplification *ethically problematic*, and why care? An obvious place to seek answers is the problematic content itself, which, as noted earlier, is an immediate source of concern. This covers all manner of misinformation and disinformation, including damaging and dangerously misleading information that can result in poor decisions and distrust. It also covers pernicious, vicious, and vile content that can lead to paranoia, radicalization, hate, bigotry, sexism, and an increasing polarization of views, opinions, and political stances. This includes, of course, aggressive, threatening, and harassing content that is not only frightening but also coercive, carrying the potential to incite hate-based violence.

Tech companies and other industry incumbents have addressed the problems of algorithmically generated feeds, reels, recommendations, and matches by taking action against problematic content. They have collectively invested billions of dollars in moderating and selectively removing content they deem unacceptable.[27] According to a 2020 report from NYU Stern School of Business, each day, over 3 million posts are flagged for potential removal by Facebook; among them, an estimated 300,000 posts are *erroneously*

---

27 *See, e.g.*, Tarasov, Katie. "Why Content Moderation Costs Billions and Is So Tricky for Facebook, Twitter, YouTube and Others." *CNBC*, February 27, 2021, https://www.cnbc.com/2021/02/27/content-moderation-on-social-media.html. Satariano, Adam, and Mike Isaac. "The Silent Partner Cleaning Up Facebook for $500 Million a Year." *New York Times*, August 31, 2021, https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html. Wagner, Kurt. "Facebook Says It Has Spent $13 Billion on Safety and Security Efforts since 2016." *Fortune*, September 21, 2021, https://fortune.com/2021/09/21/facebook-says-it-has-spent-13-billion-on-safety-and-security-efforts-since-2016/.

marked, either as a violation of standards or as benign.[28] Ethical norms dictating what content is acceptable are hardly ever clear-cut; they differ depending on culture, community, and context. Even accepting (reasonably) that some content moderation is inevitable and necessary, the practice has been controversial in some cases. For example, photos documenting war crimes and other atrocities have been censored by Facebook and other internet platforms, leading to moral outrage.[29] Beyond persistent rifts between free speech fundamentalists, on the one hand, and supporters of free speech balanced by other rights, on the other, content moderation through deletion and censorship continues to be controversial despite significant resource expenditures on them.

Content moderation and deletion, like the arcade game of "whack-a-mole," is Sisyphusian: Companies hire armies of workers to detect, identify, and scrub problematic content, and yet it quickly pops up again. Why? The answer, in our view, is rooted in the underlying algorithmic processes which, like an invisible hand, ensure its resurgence. The irony is not lost on critics that the companies paying to scrub, frequently, are the very companies responsible for the algorithmic systems they have designed, which generate clickbait[30] to hold users' attention and presence. The streams and cycles of problematic posts—as it were, the moles—will continue to appear for as long as the processes themselves, surrounding social malaise, or both are not radically addressed.

Problematic content, in our view, while important, is not the defining characteristic of problematic algorithmic amplification; instead, it stands in

---

28 Barrett, Paul M. "Who Moderates the Social Media Giants." *NYU Stern Center for Business and Human Rights* (2020): 4–5.

29 *See, e.g.*, Gillespie's discussion of the "Terror of War" photograph by Nick Ut, Facebook's moderation of the photograph, and the public and press response. In Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, 2018.

30 "Clickbait" refers to practices where headlines and other snippets of content use forward-referencing and other editorial strategies to lure, trick, or even manipulate readers into clicking, reading or otherwise engaging. Such practices are associated with low-quality content. *See* Blom, Jonas Nygaard, and Kenneth Reinecke Hansen. "Click Bait: Forward-Reference as Lure in Online News Headlines." *Journal of Pragmatics* 76 (2015): 87–100. *See also* Chakraborty, Abhijnan, et al. "Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media." *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2016.

relation to algorithmic amplification as symptoms to an underlying disease. Though it can be helpful to suppress painful symptoms, it is more lasting and decisive to address underlying causes and other sources of vulnerability. Casting instances of problematic content as one-off errors—letting in bad language through the cracks, allowing bad people to tell lies online, leading search and recommender systems to "get it wrong"—that can be handled by directly moderating it misses the point. When bad content is viewed, instead, as a *symptom*, an inevitable upshot of problematic algorithmic amplification, it draws attention to these underlying causes and, as in the case of underlying disease, a sounder, more comprehensive approach to treatment. Problematic algorithmic systems of amplification, in short, are systems that are unsuitable for the valuable purposes or tasks to which they have been put. When these systems, furthermore, nudge aside preexisting processes for performing equivalent amplifying functions, the harm is more jarring. Reducing vulnerability to individual, social, and political harm, therefore, requires that we ask two questions: First, whether algorithmic systems amplifying content are mismatched for the task their purveyors claim them to perform; and second, whether they have nudged aside historically entrenched social processes for performing these tasks.

What is a "process"?

A supporter claims that Donald Trump, actually, won the 2020 United States presidential election. When you ask why she believes this, she answers that Donald Trump said so on Twitter,[31] confirmed by many others whose Tweets appeared in her feed. Questioning this evidence, you look elsewhere for confirmation. This may take you to processes leading up to the declaration that a particular candidate has won an election. In the United States, as in other democracies, achieving the ideal of democratic elections in concrete, real-world settings means affording each individual the right and opportunity to freely cast votes in accordance with a rigorous set of procedures. A myriad rules and conventions cover each stage of an election cycle, including eligibility criteria, registration requirements, criteria for fair geographic

31 The platform is now called "X," but is referred to as "Twitter" in this paper.

representation, and protection against fraud, to name a mere handful. A myriad details define local conditions at polling stations to ensure safe and accurate voting and, later, vote counting. Where certain functions are delegated to machines, care is taken to ensure that hardware designs, systems specifications, and software code embody the required, traditional protocols for ensuring a trustworthy election.[32] Adjustments to entrenched processes may take hold over time, as existing elements are shown to be ineffective in achieving the aims and values of democratic election. For example, laws ensuring distance between pollsters, protesters, and voters might be introduced to prevent intimidation or other forms of undue influence. Finally, beyond processes for ensuring trustworthy democratic elections, we also rely on processes whereby the results of an election are announced, including the qualification required for trusted sources.

Now we introduce the idea of a social process for the purposes of this article. Briefly, a *process* refers to an ordered assemblage of conventions, rules, steps, institutions, norms, etc., guiding action, activity, and practice toward the attainment of specific ends. In the case of voting, as sketched above, the end could be described as the formation and recognition of a duly elected governing body. A process is considered successful, or *sound*, if it facilitates the fulfillment of these purported ends. It is flawed to the extent it fails[33] to do so. In the case of voting in democratic elections, polling procedures are flawed if, for example, they fail to secure confidentiality in societies where individuals are at risk of intimidation, coercion, or harm based on voting preferences. There is complexity in the ends or aims too: They may not be reducible to a simple slogan; they may or may not serve general societal welfare; they may also be parochial, serving the narrow interests of some over others; and so on. Certain behaviors and conventions in the realm of lobbying might offer a case of a political process whose ends are misaligned with broader societal goals. This paper focuses predominantly on socially

---

32 This does not mean that protocols and machines are embraced wholeheartedly by all stakeholders. Rather, we know that whether the ideals are achieved through them, and how well they are achieved, is often subject to contestation and improvement.

33 Note that "failure" can mean different things; mistaken, nonoptimal, without regard for collateral harm, etc.

valuable and just ends and identifies departures from these as problematic.

A caveat: Our aim, thus far, is not to define, in all necessary fullness, what we mean by a social process—impossible within the scope of this paper. For such a deep and complicated idea, we take comfort from the famous quote about pornography,[34] knowing the concept of *process* will be recognized through illustration, if not through formal definition. Accordingly, the aim here is to render an idea of a social process that is deeply familiar and commonly experienced, even if not explicitly recognized or named as such. Before we sketch a few more instances, we emphasize three key aspects of what we mean by social processes: 1) They are teleological, meaning they are designed to achieve an end, or set of ends; 2) like living creatures, they may change, mutate, and evolve, typically, in order to maintain efficacy against a backdrop of societal or even material (technological) change; and 3) they are prescriptive, governing behaviors implicitly, through norms and conventions, explicitly, through codes and rules, partially, or fully, through constraints embedded in technical or other material systems.[35]

Process, Aim, Alternatives

It is difficult to enumerate cases of social processes not because they are hard to find but because social life is rife with them, ranging from the trivial to the exceedingly complex, from those with little at stake to those with the highest stakes. In the workplace, processes abound, from hiring to those surrounding promotion, firing, and more. In academic life, there is a dizzying array of processes governing the review of student applicants, aspiring faculty, and journal and conference submissions. Peer review alone covers a range of differing options—single-blind, double-blind, open, secret, etc. Those of us who have experienced any of these know that they are prescriptive, sometimes highly so: They are designed around goals, such as identifying qualified applicants and accepting excellent articles. Background or contextual

---

34 *See Jacobellis v. Ohio*, 378 U.S. 184, 84 S. Ct. 1676, 12 L. Ed. 2d 793 (1964). ("I know it when I see it.")

35 A fuller discussion of the ways physical and social systems regulate behavior—and the implications for the rise of computing technologies—can be found in Lessig, Lawrence. *Code: And Other Laws of Cyberspace*. Basic Books Inc., 1999.

factors also affect the shape of a process—time, financial resources, and geographic constraints merely scratch the surface of the vast set of factors guiding behavior. We have also witnessed the complaints that ensue when processes are seen to have failed at achieving their goals. When societal or cultural prejudices, or sloppy evaluations, for example, lead to unqualified candidates being hired, when reviewers favor work from prestigious institutions instead of the most excellent, or when errors in submissions are overlooked—these failures may call for reassessment and calibration of the processes underlying these outcomes. More complex processes, such as approving new drug treatments for medical use, involve hierarchies of processes prescribing a host of component parts, including those addressing sound research practices both in the laboratory, in clinical settings, and "in the wild," whether managed by academic researchers or drug companies. These can be granular, as, for example, prescribing the random assignment of subjects to control and test groups in a trial aimed at assessing the causal effect of an intervention.[36] A myriad components and procedures constitute necessary steps toward regulatory approval, including disclosures of efficacy, risks, and ethical treatment of research subjects.[37] Why these steps, methods, practices? At the risk of stating the obvious, in the ideal scenario, they would be perceived as optimally efficacious for achieving desired ends—always open to challenge and adjustment due to contextual shifts, novel methods and technologies, and even, adjusted purposes.

Faulty component practices, tools, or conventions lead to unreliable conclusions, for example, they may yield outcomes that cannot be reproduced,[38]

---

36 The randomized control trial is sometimes characterized as holding special status compared to other causal estimation techniques—namely, it is described as a "gold standard." *See, e.g.,* Meldrum, Marcia L. "A Brief History of the Randomized Controlled Trial: From Oranges and Lemons to the Gold Standard." *Hematology/Oncology Clinics of North America* 14.4 (2000): 745–760. Even somewhat more critical accounts take up the same language, e.g., Kaptchuk, Ted J. "The Double-Blind, Randomized, Placebo-Controlled Trial: Gold Standard or Golden Calf?" *Journal of Clinical Epidemiology* 54.6 (2001): 541–549.

37 *See, e.g.,* Latour, Bruno, and Steve Woolgar. *Laboratory Life: the Construction of Scientific Fact*. Princeton University Press, 1986.

38 Examples of reproducibility issues have been identified in some psychological sciences, clinical medical research, economics, and machine-learning-based science. For corresponding studies in each of these fields, respectively, *see* Open Science Collaboration. "Estimating the Reproducibility of Psychological Science." *Science* 349.6251 (2015); Ioannidis, John P.A. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research." *Jama* 294.2 (2005): 218–228; Camerer, Colin F.,

potentially leading to unsafe drugs on the market and, ultimately, poor outcomes for health. Such issues may be attributed to poor execution of methods by individual researchers or, if persistent and pervasive, may reveal a flaw in norms, rules, conventions, methods, or tools. If that is the case, the spotlight shines on what we have called *process*. When recognized as such, problematic processes may result in targeted revisions of particular components, for example, requiring preregistration of experiments to overcome publication bias—a key cause of nonreplicable, unsound results.[39] Or, they may yield far-reaching overhauls of community practices. Whether change is achieved through systematic analysis, rhetorical calls for action, trial and error efforts, or combinations of these, for proposed adjustments to entrenched processes, the ultimate test is trustworthiness—in this instance, replicable study outcomes. Adjusted processes are justified if they are shown to be at least more successful in achieving the aims set out by respective institutional practices.[40]

Details aside, the takeaway from this section is the connection drawn between a process, that is, a structured arrangement of rules, norms, established methods and practices, standards, and guides, and the aims, or set of aims, around which it is oriented. We refer to a process as sound or trustworthy to the extent it successfully meets or promotes these aims. Although there is no reason to exclude automated algorithmic systems from this account of processes, there is also no reason to exempt them from undergoing the necessary scrutiny to establish that they meet the conditions of trust and soundness. A fair evaluation would require, further, that the automated systems in question are comparable to alternative processes for performing equivalent tasks.

Together the different strands of our discussion converge on a conception

---

et al. "Evaluating Replicability of Laboratory Experiments in Economics." Science 351.6280 (2016): 1433–1436; and Kapoor, Sayash, and Arvind Narayanan. "Leakage and the Reproducibility Crisis in Machine-Learning-Based Science." *Patterns* 4.9 (2023).

39 Nosek, Brian A., et al. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115.11 (2018): 2600–2606.

40 The aims themselves are not always utterly set in stone and may call for societal deliberation. When this happens, such shifts can ripple back to require a change in processes. The same goes for background conditions—these depend on whether a particular regulation is in place, or existing ethical codes constraining the design of drug trials, or what measurement tools and technologies researchers have at their disposal.

of ***problematic algorithmic amplification*** as formal systems that fail to achieve their purported socially valuable functions. Algorithmic systems are particularly problematic when they subsume, undermine, disrupt, erode, or replace existing processes (procedures, protocols), which may include automated systems, honed over time to achieve socially valued ends more effectively and efficiently than their algorithmic counterparts.

To signal the qualities of efficacy and efficiency, we apply the terms *sound and trustworthy*. We do so without prejudice against automated algorithmic systems, which may earn these qualities when they are shown to meet certain criteria, even when these systems replace historically entrenched processes. Accordingly, with the selective publication of ideas and content—the case that opened this article—our conception does not support prior practices where the full power falls to a privileged few to decide, without account, what is published or earns airtime, etc. Trustworthy processes require reflective assessment against a society's valued ends and, likely, will continue being refined and improved in competition with reasonable alternatives, algorithmic or otherwise.

Returning to the supporter who is affirmed in her belief that Trump won the 2020 election based on an overwhelming proportion of Tweets in her Twitter feed. One might ask why Twitter's selective amplification does not qualify as a social process as we have described it, alongside the more analog and long-standing processes that, traditionally, have established and proclaimed the winners of political elections. This question is legitimate and sits at the crux of our substantive thesis. Yes, the sociotechnical systems of algorithmic amplification can be conceived of as types of processes. In this instance, the reasons for favoring entrenched traditional processes over Twitter's algorithmic amplification, however, are rooted in their respective aims (and their performance vis-a-vis their aims). In the entrenched process, these aims are to achieve and proclaim democratically elected leaders and they stand or fall on their capacities to do so. When they fail, or are believed to perform suboptimally, they become targets of adjustment or even replacement.[41] Approaches to evaluation, whether empirical, historical, analytical,

---

41 Conflicting views on whether voter identification rules discriminate against racial and ethnic

statistical, or other, are informed by fundamental commitments to democracy and legitimate approaches to governance, given popular expression in maxims, such as, "of the people, by the people, for the people." This is a far cry from the litmus test of Twitter's algorithmic systems for constructing individual feeds (to the extent we are able to know these). Twitter is a privately controlled corporate platform whose primary commitment is to building profit for its shareholders by directing user attention toward paying advertisers' messages on the platform. To serve these aims, Twitter's "algorithm"[42] displays content likely to capture user attention and promote engagement.

Different algorithmic systems may be more or less appropriate when appraised as part of different processes (and in light of different aims). For example, Reddit's algorithms rank posts based on the numbers of "up-votes" and "down-votes" they receive from users.[43] Members of subcommunities ("subreddits"), relying on an algorithm that elevates posts based on numbers of up-votes, might extol these as "democratic" and might find them useful for locating memes, stories, art, and other cultural content that are generally resonant, inspiring, or uplifting. These algorithmic ranking mechanisms determined by up-votes, however, are not necessarily appropriate for all ends; for example, scientific findings about the risks of vaccines, the casualty count in a distant war, or the threat level of an impending natural disaster. A voting-based mechanism that tracks popularity does not guarantee that true, reliable findings will make their way to the top of a Reddit feed. If the *purpose*

---

minorities may lead to practical changes, one way or another. Though proponents claim that they protect against voter fraud, critics point to their chilling effects on participation (in particular, the participation of racial and ethnic minorities) as evidence that they impede democratic goals. As legislatures and courts enact policy changes, it is no guarantee that more successful norms and practices take hold. Cases such as this are vulnerable to problematic amplification. See, e.g., Bronner, Ethan. "Partisan Rifts Hinder Efforts to Improve U.S. Voting System." *The New York Times*, 2012, https://www.nytimes.com/2012/08/01/us/voting-systems-plagues-go-far-beyond-identification.html.

42 The term "algorithm" is used capaciously. For a discussion of the uses and limits of this terminology, *see* Lum, Kristian, and Lazovich, Tomo. "The Myth of the Algorithm: A System-Level View of Algorithmic Amplification." *Knight First Amendment Institute*, September 13, 2023, https://knightcolumbia.org/content/the-myth-of-the-algorithm-a-system-level-view-of-algorithmic-amplification.

43 Reddit allows users to choose from a number of different ranking algorithms, including those called "top," "hot," and "new." Depending on the choice of ranking algorithm, voting may not be the only factor used. For example, the amount of time elapsed since the post's publication date also factors into Reddit's ranking. For a fuller discussion of its algorithm, *see* Munroe, Randall. "Reddit's New Comment Sorting System." *Reddit Blog*, October 15, 2009, http://redditblog.blogspot.com/2009/10/reddits-new-comment-sorting-system.html.

is to convey trustworthy information, a popularity-based algorithmic system for elevating and demoting certain posts may not, systematically, align with the purpose of highlighting the most trustworthy information.

THE CASE OF EPISTEMIC PROCESSES

FOR CENTURIES, THE PROCESS for disseminating and amplifying information has co-evolved with the process for vetting claims as truthful and trustworthy. People have collectively invested in institutions and processes aimed at verifying and underwriting claims as factually truthful. They have designed and vetted institutions and processes as trustworthy, in order to address and resolve disagreements over factual claims. But these arrangements, and our faith in them, are fraying. Curatorial practices on the web are driving a wedge between truth claims that are algorithmically amplified and those that have been vetted through sound, trustworthy processes. This section focuses on the family of threats that plague epistemic processes as a test case for our account of problematic algorithmic amplification, not because they are the only threats or processes we should be concerned about, but because they are particularly potent and wide-reaching.

Digital platforms, including the web itself, have raised the stakes for epistemology, the branch of philosophical inquiry into the nature of knowledge—how we come to know, and the validity of knowledge claims. Since ancient times, and through the present day, debates have raged over whether knowledge requires direct experience with one's senses, rational analysis from premises to conclusion, deep and deliberative reflection about the conceptual world, exposure of belief claims to counterattack by adversaries,[44]

---

44 Writing at a time when the church's teachings clashed with scientific findings, Mill (1859) concluded that public questioning and discussion—even of unsound claims—was crucial to arrive at reliable conclusions. To that end, freedom from censorship and freedom of thought became bedrock norms for a trustworthy epistemic process. *See* Mill, John Stuart. *On Liberty and Other Essays*. Oxford University Press, 1998. Out of this intellectual tradition, legal protections for public speech have taken root in the United States, and First-Amendment scholarship has wrestled with questions around the coverage and limits of these protections. Working, specifically, on social media amplification, Miller (2021) pointed out that amplifying speech to large audiences can distort the

and more. It takes only a moment's reflection to understand that if "seeing is believing," then believing only what one directly experiences with one's senses would shrink the universe of belief to almost nothing. A moment longer and we grasp that almost all of what we know, or claim to know, comes not from direct experience but from the testimony of others, including what we learn from family, community, school, books, newspapers, TV, and more. The degree to which we depend on others, on sources outside of ourselves, for almost all that we know, explains why the stakes are so high. Digital platforms and websites, increasingly, have joined and taken over as dominant sources and curators of what people believe and claim as knowledge. Herein lies one of the serious risks of problematic algorithmic amplification: Inundated by information vying, algorithmically, for our attention and credence and calling, algorithmically, for our commitment, action, and engagement, we are frequently at a loss as to what to trust, whom to trust, and why.

These thorny issues have been pondered for centuries. In the Socratic dialogues, Plato relentlessly attacks his archenemies, the *sophists*,[45] arguing that their clever rhetoric amounts to a "sort of conjectural or apparent knowledge only of all things, which is not the truth."[46] Their performances, paid for by wealthy patrons, misled, even corrupted their audiences. To this day, sophists and sophistry are associated with false and deceptive rhetoric,[47] unreliable means toward untrustworthy knowledge claims. By contrast, Plato's Socrates is depicted as a *philosopher*, humble in acknowledging his ignorance. A philosopher is not an oracle but a general seeker of knowledge,[48] attained through careful, systematic argumentation and deliberately

---

"marketplace of ideas," a long-standing analogy in First-Amendment doctrine and political theory that conceives of public discourse as a competitive marketplace where, ideally, conditions favor reliable, trustworthy, justified, true beliefs. Miller questions the laissez-faire approach to free speech protections. *See* Miller, Erin L. "Amplified Speech." *Cardozo L. Rev*. 43 (2021): 1. *See also* Blocher, Joseph. "Free Speech and Justified True Belief." *Harv. L. Rev.* 133 (2019): 439.

45 Sophists were people who taught and lectured in ancient Greece. *See, e.g.*, Taylor, C.C.W. and Mi-Kyoung Lee, "The Sophists," *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2020/entries/sophists/.

46 Plato, *Sophist*, translated by Benjamin Jowett, http://classics.mit.edu/Plato/sophist.html.

47 *See* the definition of sophistry. E.g., Miriam-Webster defines sophistry as "subtly deceptive reasoning or argumentation," https://www.merriam-webster.com/dictionary/sophistry.

48 His famous claim inspired the modern quote, "I know that I know nothing," a radical departure from the sophist's way. For more in-depth discussion, *see* Vlastos, Gregory. "Socrates'

exposing claims to rebuttal, challenge, and test. Known as the Socratic method, the approach comprises a process of back-and-forth assertions among interlocutors who ask and challenge one another in a sequence of questions aimed at flushing out contradiction and, ultimately, the refinement of an idea. For Plato, argument through deductive, conceptual, and theoretical reasoning was the most reliable process for acquiring knowledge, superior even to direct perception.[49] Though not recommending a return to Plato's epistemology, we see these ancient writings as a reminder that truth is not all there is to knowledge. As important are the processes we rely on for achieving it.

We use the term *epistemic process* (or *knowledge process*) to capture, in broad strokes, the structured practices aimed at achieving knowledge, understanding, expertise, or, generally, at justifying truth claims. Analogous to the processes for voting and those for identifying safe medical treatments, epistemic processes are rich assemblages of procedures, methods, actions, rules, norms, conventions, institutions, etc., whose aim, in this case, is knowledge. Different knowledge-seeking activities may be relied on for a vast array of knowledge types, ranging from questions about a train schedule to highly specialized questions, such as the causes of brain cancer, the origins of the universe, the state of a faraway war, or the costs and benefits of childhood vaccinations.

A child in a library completing a history project about, say, the Middle East, might pull a textbook or encyclopedia from the library stacks or might turn to Wikipedia or Google Search. Although these sources may seem roughly interchangeable, the epistemic processes they rely on are not. Reputable textbooks or hard-bound encyclopedias conform to an assortment of norms governing the selection of topics and sources, expert authorship, primary source engagement, credible citations, editorial scrutiny, etc.[50] That they sit on library shelves, furthermore, implies that they have weathered

---

Disavowal of Knowledge." *The Philosophical Quarterly* (1950–) 35.138 (1985): 1–31.

49 Plato, *The Republic*, Book VII, https://www.gutenberg.org/cache/epub/1497/pg1497-images.html#link2H_4_0009.

50 Schroeder, Milton R., and Mary M. Schroeder. "The New Encyclopaedia Britannica: All Human Knowledge." *ABAJ* 60 (1974): 711.

systematic vetting by librarians. Wikipedia, which may seem equivalent to hardcopy encyclopedias in a different medium, conforms with markedly different norms, relying on networks of volunteers who contribute, edit, and moderate pages.[51] Whether Wikipedia earns people's trust depends on a number of factors, such as whether independent processes and trusted sources corroborate its entries, and whether people approve of the norms and rules (in our terms, the processes) that Wikipedia supports for evaluating the merits of a given entry (including familiar warnings when these processes have not been followed).[52]

When students turn to Google Search, they are relying on yet another process—a complex set of Google's famous and secret indexing and ranking algorithms, which place on their screen a list of sources, including some that have paid for placement. Google Scholar, which might, arguably, stand in for the vetting processes of librarians, similarly ranks and recommends sources of information according to an opaque algorithmic system, leaving to Google a host of evaluative decisions about the relative importance of citation counts, download counts, journal, title, and text for establishing credibility and for determining relevance.[53]

What makes an epistemic process successful? What supports the conclusion that a particular assemblage of norms, practices, methods, procedures, etc. is a reliable pathway to knowledge? Philosopher John Hardwig pointed to trust as an essential component. For Hardwig, people are epistemically tied to one another through trust and testimony, and thus, "the trustworthiness of members of epistemic communities is the ultimate foundation for much

---

51 Although a fuller discussion of Wikipedia is outside the scope of this article, interested readers may learn more from Joseph Reagle's excellent book on the history and philosophy of Wikipedia. Reagle, Joseph. *Good Faith Collaboration: The Culture of Wikipedia*. History and Foundations of Information Science. The MIT Press, 2010.

52 Transparency is one such norm that, in conjunction with other practices, may give people reason to trust a process. Simon (2010) pointed to transparency as a "fundamental requirement" for trustworthy knowledge generation online, using Wikipedia as an extended case. *See* Simon, Judith. "The Entanglement of Trust and Knowledge on the Web." *Ethics and Information Technology* 12.4 (2010): 343–355.

53 The choices made by Google Scholar also hold profound relevance for the direction of scholarly research and the integrity of evaluations. For a fuller discussion of how they both disrupt and threaten trusted processes, *see* Goldenfein, Jake, and Daniel S. Griffin. "Google Scholar—Platforming the Scholarly Economy." *Internet Policy Review* 11.3 (2022).

of our knowledge."[54] Epistemic processes organize and structure epistemic communities and, depending on their characteristics and qualities, may give us more or less reason to place our trust in them. Trusting a process means placing faith in its conclusions; without trust, the process loses its capacity to instill new beliefs. Trust alone, however, makes us vulnerable and exposes us to risk; it does not ensure that our beliefs are sound. Accordingly, a process must also be *trustworthy*. As we have described in this paper, processes exhibit trustworthiness, according to our definition, if they follow norms, conventions, and rules that, generally, are successful in serving their aims. In the case of epistemic processes, this means that they generate reliable knowledge claims. Notice that trustworthiness, alone, without trust, absent communal or societal buy-in, may also be insufficient for knowledge.

Returning to our hypothetical scenario with a further twist, we find that our wayward student, having procrastinated too long on his research project, resorts to his favorite social media application (interchangeably, Instagram, Facebook, Twitter, Snapchat, TikTok, or YouTube). Scrolling through endless reels and feeds[55] of content, he sees posts conveying not only information about his friends and family but also news about an ongoing war in the Middle East. Seeking more information, he expands a post and reads updates about the state of things. In doing so, he is relying, perhaps inadvertently, on a social media company's algorithmic recommendation as a component in his epistemic discovery. But the communicative process detailing the production and direction of information about war in the Middle East to the student's phone is not geared, solely, to a sound and trustworthy account of events. The secret algorithmic brews that shape what stories are served to

---

54 Hardwig, John. "The Role of Trust in Knowledge." *The Journal of Philosophy* 88.12 (1991): 693–708. Page 694. Also relevant is Latour and Woolgar (1979).

55 From the feature's introduction in 2006 until 2022, Facebook used the name "News Feed" for its page that centrally ranks, recommends, and displays posts and updates from friends. The feature was rebranded in 2022 under the shorter name, "Feed"—perhaps implicitly disavowing its use as a quality source of news and perhaps inadvertently highlighting critics' concerns about habit-forming and dependence. For information and context around the history of this feature and its name, *see* Clark, Mitchell. "Facebook Rebrands News Feed after More Than 15 Years." *The Verge*, February 15, 2022, https://www.theverge.com/2022/2/15/22935080/facebook-meta-news-feed-renaming-branding-political-content-misinformation. *See also* Manjoo, Farhad. "Can Facebook Fix Its Own Worst Bug." *The New York Times Magazine*, April 25, 2017, https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html.

which readers are optimized for user attention and engagement more than the reliability and quality of the producers of news stories or the knowledge and expertise of sources.[56] The same risk holds for the recommendations of search engines, which may also be distorted by the exploits of those wishing to be found[57] as well as the presumed interests of searchers and seekers.

If information is recommended because it attracts attention—measured via clicks, taps, scrolls, and time spent on a site—no one should be surprised that these systems disserve the aims of knowledge and understanding; soundness does not, or does not fully, figure into the process. One should also not be surprised that contradictory or inconsistent claims are amplified simultaneously, by the same platform, because both capture attention. The posts that engagement-driven algorithms tend to elevate and amplify—short, simple, controversial, vitriolic, humorous—may well be true, but if true, only accidentally. In short, if knowledge is the aim, these algorithmic systems of amplification are not trustworthy as epistemic processes.

Often contested and controversial, the domain of news reporting has experienced significant upheaval from the growing dominance of digital media and, even more so, incursions by online platforms.[58] Problematic algorithmic amplification poses a particularly acute challenge, undermining processes that have held sway and have supported the epistemic standing of traditional news outlets (e.g., The Wall Street Journal, The New York Times, Le Monde, The Hindu, Financial Times, The Asahi Shimbun, etc.).

---

56 For years, Facebook's algorithmically ranked feeds were tuned to favor posts that received "angry" reactions, over those that did not. Though down-voting a post or comment on other platforms like Reddit would negatively impact its ranking on others' feeds, an "angry" reaction on Facebook did the opposite: It helped amplify the post, offering it a boost in ranking five times greater than that offered by a "like." *See* Merrill, Jeremy B., and Will Oremus. "Five Points for Anger, One for a 'Like': How Facebook's Formula Fostered Rage and Misinformation." *The Washington Post*, October 26, 2021, https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/. For a general discussion of optimization and its potential ethical pitfalls, *see* Laufer, Benjamin, Thomas Gilbert, and Helen Nissenbaum. "Optimization's Neglected Normative Commitments." *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (2023).

57 Recognized already in Introna, Lucas D., and Helen Nissenbaum. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16.3 (2000): 169–185.

58 *See* Nielsen, Rasmus Kleis, and Richard Fletcher. "Democratic Creative Destruction? The Effect of a Changing Media Landscape on Democracy." *Social Media and Democracy: The State of the Field, Prospects for Reform* (2020): 139–162.

Trustworthy news reporting demands adherence to sound epistemic processes, including an array of norms, conventions, and assumptions around key practices of news publication. These include selecting and prioritizing stories, reliable and impartial reporting of facts, clarity of language, and more. News outlets explicitly endorse their commitments through public statements of editorial policies,[59] and in practice, their editorial leadership may implement oversight and enforcement mechanisms. As an illustration, we highlight attribution. Not limited, of course, to the reporting of news, attribution is a centuries-old practice of linking assertions of fact or opinion to the testimonials of others—to sources. How effectively attribution validates these assertions depends on the authority, credibility, expertise, or perspective of sources and the viability of access by the writer, among other factors. In academic research publication, for example, citation is one of the key modes of attribution and is further governed by a host of granular norms, veering to the banal (e.g., whether to list first or last name first, what information to italicize, etc.) Attribution not only serves to validate assertions by citing sources, but it also serves as a means of engendering research integrity,[60] acknowledging priority, and crediting others for innovative ideas.[61]

In news reporting, too, attribution is critical. At the highest layer, disclosing the identity of the reporter not only provides credit but also connects a claim to an individual who is responsible for the article and its account of events. One layer down, good reporting will cite and quote other authors and sources, ideally a plurality of them, identified either by name or by role and frequently, both. The Associated Press, for example, publishes guidelines

---

59 *See, e.g.*, "Newsroom Standards & Ethics." *The Wall Street Journal*, https://newsliteracy.wsj.com/standards-and-ethics/. "Standards and Ethics." *The New York Times*, https://www.nytco.com/company/standards-ethics/. "Groupe Le Monde's Code of Ethics and Professional Conduct." *Le Monde*, 2022, https://www.lemonde.fr/en/about-us/article/2022/04/06/groupe-le-monde-s-code-of-ethics-and-professional-conduct_5979823_115.html. "Living Our Values: Code of Editorial Values." *The Hindu*, 2011, https://www.thehindu.com/news/national/living-our-values-code-of-editorial-values/article1715043.ece.

60 McNutt, Marcia K., et al. "Transparency in Authors' Contributions and Responsibilities to Promote Integrity in Scientific Publication." *Proceedings of the National Academy of Sciences* 115.11 (2018): 2557–2560.

61 Nissenbaum, Helen. "New Research Norms for a New Medium." *The Commodification of Information* (2002): 433–457.

for reporting on various sources under various circumstances, including cases where it is necessary to keep a source anonymous—these cases require meeting additional standards such as managerial approval, additional corroborating sources, and explanations in the article detailing why the source is credible and their reason for requesting anonymity.[62] These norm-driven practices are not arbitrary. As Hardwig might affirm, they contribute to the integrity of a story, and establish both its soundness, impartiality, and, ultimately, its trustworthiness. Finally, and perhaps most crucially, ideal attribution practices generate transparent lines of accountability from publishers, through authors, to sources.

Analogous practices taking hold on the internet—on webpages, social media sites, blogs, apps, etc.—loosely resemble long-standing attribution practices. Although many webpages are not associated with named authors, hyperlinks may be used to source claims. In fact, the network structure that these links constitute inspired automated internet search algorithms, such as Google Search, to locate useful and authoritative information.[63] But the policies and practices of online platforms have not, uniformly, embodied trustworthy attribution practices, eventually contributing to degraded content. For example, Twitter has apparently penalized posts with external links[64]

---

62 Associated Press. "The Associated Press Statement of News Values and Principles," https://www.ap.org/about/news-values-and-principles/downloads/ap-news-values-and-principles.pdf.

63 These algorithms, often, aim to identify "authoritative" sources using a graph of hyperlinks. *See* Kleinberg, Jon M. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM (JACM)* 46.5 (1999): 604–632.

64 Experts in media strategy have suggested that Twitter penalizes posts with external links, as noted in Kirshner, Alex. "Twitter Was for News." *Slate*, October 5, 2023, https://slate.com/technology/2023/10/elon-musk-x-twitter-news-links-headlines-why.html. The article also describes how Twitter in 2023 removed headlines and other contextual information that automatically appeared in Tweets with hyperlinks. In an article on LinkedIn, Bernhardt suggests "Links Hurt." Bernhardt, Sarah Larsson. "Twitter Revealed Its New Algorithm to the World. This Is What We Know So Far." *LinkedIn*, April 3, 2023, https://www.linkedin.com/pulse/twitter-revealed-its-new-algorithm-world-what-we-know-sarah. Empirical studies support the hypothesis that external hyperlinks are penalized on Twitter's algorithmic feed compared to a chronological feed. In an "agent-based" experiment with data from eight accounts created by the researchers to emulate real users, Bandy and Diakopoulos (2021) find that the algorithmic timeline exposes agents to about half as many links as the chronological timeline. Bandy, Jack, and Nicholas Diakopoulos. "Curating Quality? How Twitter's Timeline Algorithm Treats Different Types of News." *Social Media + Society* 7.3 (2021). Milli et al. (2023), using a randomized control trial of real users, find the same effect. Milli et al. (2023): Appendix E.1.

and blocked links to competing platforms,[65] presumably to reduce user traffic and attention to sources outside of Twitter. This means those who aim to follow nascent norms of attribution (academics, journalists, and others) to signal the trustworthiness of their posts are penalized. Instead, Twitter's algorithms reward those who "screenshot" catchy visuals or headlines, cut out detailed text or descriptions of method, omit information about the publisher, and remove hyperlinks. When scientific plots and visualizations spread on Twitter and similar platforms now, readers are deprived of the means to locate sources and, ultimately, are less equipped to evaluate them. Platforms that apply algorithmic systems, tuned to user attention and engagement, undermine their own trustworthiness as an epistemic source while simultaneously sowing suspicion of long-trusted and trustworthy sources, including family members, norm-abiding news outlets, or encyclopedias.

Experts have noted that misinformation is endemic to major internet platforms, including YouTube,[66] Facebook,[67] Twitter,[68] and elsewhere.[69] Algorithmic recommender systems, which utilize historical user behavior to profile and draw inferences about user preference, can exacerbate misinformation by personalizing their recommendations.[70] For example, in Hussein

---

65 Twitter has blocked links to rivaling platforms including Meta's Threads and Mastodon. *See* Roth, Emma. "Twitter Abruptly Bans All Links to Instagram, Mastodon, and Other Competitors." *The Verge*, December 18, 2022, https://www.theverge.com/2022/12/18/23515221/twitter-bans-links-instagram-mastodon-competitors. *See also* Perez, Sarah. "Twitter Blocks Links to Rival Threads, While CEO Downplays Reports of Traffic Decline." *TechCrunch*, July 11, 2023, https://tcrn.ch/46GHcMP.

66 Hussein, Eslam, Prerna Juneja, and Tanushree Mitra. "Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube." *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW1 Article 48 (2020): 1–27.

67 Del Vicario, Michela, et al. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences* 113.3 (2016): 554–559.

68 *See* Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359.6380 (2018): 1146–1151. For an example where Twitter misinformation was especially prominent, *see* Suarez-Lledo, Victor, and Javier Alvarez-Galvez. "Prevalence of Health Misinformation on Social Media: Systematic Review." *Journal of Medical Internet Research* 23.1 (2021).

69 *See, e.g.*, Allen, Jeff. "Misinformation Amplification Analysis and Tracking Dashboard." *Integrity Institute*, October 13, 2022, https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard.

70 Some have argued that personalized "filter bubbles" can feed people information they want to see, skewing their understanding of facts and events. *See* Pariser, Eli. *The Filter Bubble: How the*

et al.'s (2020) study of YouTube's personalized search and recommendation algorithms, they found that personalization based on watch history increases the prevalence of misinformation appearing in search results and recommended videos.[71] Empirical findings, such as these, demonstrate the ongoing prevalence of fake or false information online but, as we have argued throughout this article, do not address the root causes of the epistemic risks we confront. Our focus should be on problematic algorithmic systems insidiously taking the place of processes that we have relied on (sometimes unconsciously and implicitly) for pointing to knowledge and truth. Experimental hypothesis testing, investigative journalistic standards, peer review, trial proceedings, indexing and references, encyclopedia publishing norms, classroom teaching methods, mathematical proof—these all constitute processes and procedures, typically employed with the aim of generating sound and practical knowledge. Whether explicit and formalized or implicit and taken for granted, the processes we rely upon are fundamental to the proper functioning of legal, industrial, and social systems.

The endpoint of this downward trend is that societies with weakened, fragmented epistemic processes might lose the capacity to distinguish between reliable reporting and disinformation and fail to find common ground among believers of opposing facts—a modern-day Tower of Babel.

CONCLUSION

THE CENTRAL ARGUMENT of this paper is that algorithmic amplification is problematic not, primarily, because the content it amplifies is problematic. Instead, algorithmic amplification is problematic

New Personalized Web Is Changing What We Read and How We Think. Penguin Books, 2012. People's understanding can be more severely threatened when external sources and voices are actively discredited, in what Nguyen (2020) defines as an "echo chamber." Nguyen, C. Thi. "Echo Chambers and Epistemic Bubbles." *Episteme* 17.2 (2020): 142.

71 Only personalization based on watch history—not demographics—had a measurable effect on misinformation recommendations. Analysis further revealed a "filter bubble effect." *See* Hussein et al. (2020).

because, like an invasive species, it chokes out trustworthy processes that we have relied on for guiding valued societal practices and for selecting, elevating, and amplifying content. Problematic content is a symptom, only partially addressed when disassociated from its underlying causes. Sound processes, continually fine-tuned to illuminate pathways toward valued and valuable societal ends (e.g., from reliable procedures to trustworthy knowledge claims) have inoculated us against processes devised by stakeholders to promote their own or other parochial interests (e.g., cost-effectiveness or profitability). We have argued that in an inexorable progression toward digital life, our societies have become increasingly reliant on computationally encoded algorithmic systems, such as those governing the distribution of content on digital media and platforms, managing voting machines, and automating a host of life-critical decisions, to name a few. We are not advocating for a reversal to an old order that privileges exclusive elites. Nor are we opposed to automated algorithmic systems being embedded in sound processes. Instead, we call attention to what is at stake when processes tuned to the fulfillment of aims and values held in common are nudged aside by algorithmic systems tuned (sometimes stealthily) to different, even incompatible ends.

As the internet increasingly mediates all aspects of our lives, we are likely to have encountered some form or another of problematic content—clickbait, misinformation, extremist ideology, bigotry, among other varieties. This content may have made us feel uncomfortable, distressed, angry, even disgusted. We have argued here that the stakes of problematic algorithmic amplification extend well beyond these experiences. Algorithmic systems have demonstrated astonishing capabilities and have inspired hopes for more informed citizens and more robust democratic institutions. In this optimism, we may not notice that when algorithms nudge aside legacy processes in the name of efficiency and scale, they leave us vulnerable to dogmatic machine oracles or charming present-day sophists. Bad content, we have argued, is the canary in the coal mine, warning of a deeper problem that will not be cured by more diligent moderation or more solicitous recommendation.

A POSTSCRIPT CONCERNING GENERATIVE AI

While working on this article, generative AI burst onto the public scene and, particularly with the release of ChatGPT, sparked widespread fascination. In print and informal discussions, experts and nonexperts have expressed awe over its capacities to reason, converse, and generate knowledge and information, while pundits and promoters are already predicting revolutionary change in many walks of life, including healthcare, law, finance, and education.[72] At the same time, many commentators have warned of dangers and threats, including threats to the livelihoods of creative artists, libelous and defamatory assertions, privacy violations, and more. We were immediately drawn to the many reports of errors, "hallucinations," and falsehoods, among other strange and toxic behaviors.[73] Based on our findings, we know that the immediate reaction—to delete or correct falsehood—is to miss the most insidious problem, namely, the gradual extinction of processes that have soundly guided us toward knowledge and sound decision-making.

Our account of problematic algorithmic amplification holds equally urgent lessons for generative AI. In the foreseeable future, as we welcome AI-powered tools across social domains, it will remain necessary to fact-check, corroborate, justify, and moderate the content that these tools produce. This will not be possible unless we maintain sound and strong epistemic processes, robustly sustained by social systems and institutions. Erosion of these processes through diminished support or simple neglect

---

72 For example, Kevin Roose, a columnist with *The New York Times*, enthusiastically supports the use of ChatGPT in classrooms and calls on teachers to learn how to utilize it for the sake of students and for their own. Roose, Kevin. "Don't Ban ChatGPT in Schools. Teach With It." *The New York Times*, January 12, 2023, https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html.

73 *See, e.g.*, Hunt, Elle. "Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter." The Guardian, March 24, 2016, https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter. Wiggers, Kyle. "Researchers Discover a Way to Make ChatGPT Consistently Toxic." *TechCrunch*, April 12, 2023, https://techcrunch.com/2023/04/12/researchers-discover-a-way-to-make-chatgpt-consistently-toxic/. Walsch, Toby. "Gaslighting, Love Bombing and Narcissism: Why Is Microsoft's Bing AI So Unhinged?" *The Conversation*, February 17, 2023, https://theconversation.com/gaslighting-love-bombing-and-narcissism-why-is-microsofts-bing-ai-so-unhinged-200164. Needleman, Sarah. "Microsoft Defends New Bing after AI Chatbot Offers Unhinged Responses." *Wall Street Journal*, February 17, 2023, https://www.wsj.com/tech/ai/microsoft-defends-new-bing-says-ai-upgrade-is-work-in-progress-3447074d.

leaves us terribly exposed. When we rely on generative AI at the expense of preexisting processes and allow these sound processes to atrophy, the logical end point of this displacement is a society with unimaginable vulnerability—the loss and weakening of other standards against which to check the veracity or soundness of AI assertions, no counterpoints for students to critically assess ChatGPT's hallucinations, no independent sources beyond the problematic algorithmic systems themselves, nowhere to turn for answers.

About the Authors

Benjamin Laufer is a Ph.D. Candidate at Cornell Tech. He studies the values and politics embedded in technological systems, particularly those deployed in high-impact, high-complexity domains in the public realm. He is a doctoral fellow at the Digital Life Initiative and an affiliate of the Artificial Intelligence, Policy, and Practice group at Cornell University. He is advised by Helen Nissenbaum and Jon Kleinberg. He holds a B.S.E. (cum laude) in operations research and financial engineering from Princeton University.

Helen Nissenbaum is the Andrew H. and Ann R. Tisch Professor of Information Science and the founding director of the Digital Life Initiative at Cornell Tech. Her work focuses on ethical and political implications of digital technologies on issues such as privacy, bias in digital systems, trust online, ethics in design, and accountability in computational and algorithmic systems. Grants from the National Science Foundation, Air Force Office of Scientific Research, the U.S. Department of Health and Human Services Office of the National Coordinator, McArthur Foundation, Defense Advanced Research Projects Agency (DARPA), and the National Security Agency have supported her research. Recipient of the 2014 Barwise Prize of the American Philosophical Association and the International Association of Computing and Philosophy Covey Award for computing, ethics, and philosophy, Nissenbaum has contributed to privacy-enhancing free software, TrackMeNot (protecting against profiling based on web searches) and AdNauseam (protecting against profiling based on ad clicks). She holds a Ph.D. in philosophy from Stanford University and a B.A. (Hons) in philosophy and mathematics from the University of the Witwatersrand, South Africa.

Acknowledgments

This work is supported by a grant from the John D. and Catherine T. MacArthur Foundation and a SaTC NSF grant CNS-1704527. Ben Laufer is additionally supported by a Bowers CIS-LinkedIn PhD Fellowship and a Digital Life Initiative Doctoral Fellowship. We thank the members of the Digital Life Initiative (DLI) at Cornell Tech, the AI, Policy, and Practice (AIPP) group at Cornell University, and the attendees, participants, and organizers of the Symposium on Algorithmic Amplification and Society at the Knight First Amendment Institute at Columbia University. We acknowledge the following people for their comments and suggestions: Robyn Caplan, Sarah Cen, Stephanie May Chan, Amelie Ortiz De Leon, Katy Glenn Bass, Seth Lazar, Smitha Milli, Mor Naaman, Arvind Narayanan, Judith Simon, Mia Speier, Daniel Susser, and Maya Von Ziegesar.

Cite as: Benjamin Laufer and Helen Nissenbaum, *Algorithmic Displacement of Social Trust*, 23-12 Knight First Amend. Inst. (Nov. 29, 2023), https://knightcolumbia.org/content/algorithmic-displacement-of-social-trust [https://perma.cc/2TPM-JGWX].

About the Knight First Amendment Institute

The Knight First Amendment Institute at Columbia University defends the freedoms of speech and the press in the digital age through strategic litigation, research, and public education. It promotes a system of free expression that is open and inclusive, that broadens and elevates public discourse, and that fosters creativity, accountability, and effective self-government.

knightcolumbia.org

Design: Point Five
Cover illustration: Emilie Flamme

References

Agnone, Jon. "Amplifying Public Opinion: The Policy Impact of the US Environmental Movement." *Social Forces* 85.4 (2007): 1593–1620.

Allen, Jeff. "Misinformation Amplification Analysis and Tracking Dashboard." *Integrity Institute*, October 13, 2022, https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard.

Ashby, William R. *An Introduction to Cybernetics*. Chapman & Hall, 1956.

Associated Press. "The Associated Press Statement of News Values and Principles," https://www.ap.org/about/news-values-and-principles/downloads/ap-news-values-and-principles.pdf.

Bandy, Jack, and Nicholas Diakopoulos. "Curating Quality? How Twitter's Timeline Algorithm Treats Different Types of News." *Social Media + Society* 7.3 (2021).

Barrett, Paul M. "Who Moderates the Social Media Giants." *NYU Stern Center for Business and Human Rights* (2020): 4–5.

Beckett, Louis, and Julia Carrie Wong, "The Misinformation Media Machine Amplifying Trump's Election Lies." *The Guardian*, November 10, 2020, https://www.theguardian.com/us-news/2020/nov/10/donald-trump-us-election-misinformation-media.

Benjamin, Ruha. *Race after Technology*. Polity Press, 2019.

Bernhardt, Sarah Larsson. "Twitter Revealed Its New Algorithm to the World. This Is What We Know So Far." *LinkedIn*, April 3, 2023, https://www.linkedin.com/pulse/twitter-revealed-its-new-algorithm-world-what-we-know-sarah.

Bermeo, Nancy. "On Democratic Backsliding." *J. Democracy* 27.1 (2016).

Blocher, Joseph. "Free Speech and Justified True Belief." *Harv. L. Rev.* 133 (2019): 439.

Blom, Jonas Nygaard, and Kenneth Reinecke Hansen. "Click Bait: Forward-Reference as Lure in Online News Headlines." *Journal of Pragmatics* 76 (2015): 87–100.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. "Cross-Country Trends in Affective Polarization." *Review of Economics and Statistics* (2022): 1–60.

Bronner, Ethan. "Partisan Rifts Hinder Efforts to Improve U.S. Voting System." *The New York Times*, 2012, https://www.nytimes.com/2012/08/01/us/voting-systems-plagues-go-far-beyond-identification.html.

Camerer, Colin F., et al. "Evaluating Replicability of Laboratory Experiments in Economics." Science 351.6280 (2016): 1433–1436.

Cen, Sarah H., Aleksander Madry, and Devavrat Shah. "A User-Driven Framework for Regulating and Auditing Social Media." *arXiv preprint arXiv:2304.10525* (2023).

Chakraborty, Abhijnan, et al. "Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media." *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2016.

Clark, Mitchell. "Facebook Rebrands News Feed after More Than 15 Years." *The Verge*, February 15, 2022, https://www.theverge.com/2022/2/15/22935080/facebook-meta-news-feed-renaming-branding-political-content-misinformation.

Cranor, Lorrie F. "A Framework for Reasoning about the Human in the Loop." *Proceedings of the 1st Conference on Usability, Psychology, and Security* (2008).

De Zeeuw, Daniel, et al. "Tracing Normiefication: A Cross-Platform Analysis of the QAnon Conspiracy Theory." *First Monday* 25.11 (2020).
Del Vicario, Michela, et al. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences* 113.3 (2016): 554–559.
Donovan, Joan, and danah boyd. "Stop the Presses? Moving from Strategic Silence to Strategic Amplification in a Networked Media Ecosystem." *American Behavioral Scientist* 65.2 (2021): 333–350.
Eckles, Dean. "Algorithmic Transparency and Assessing Effects of Algorithmic Ranking." *Testimony Before the Senate Subcommittee on Communications, Media, and Broadband*, December 9, 2021, https://www.commerce.senate.gov/services/files/62102355-DC26-4909-BF90-8FB068145F18.
Etymonline, "amplify (v.)," https://www.etymonline.com/word/amplify.
Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* Yale University Press, 2018.
Gillespie, Tarleton. "The Relevance of Algorithms." *Media Technologies: Essays on Communication, Materiality, and Society* (2014): 167.
Goldenfein, Jake, and Daniel S. Griffin. "Google Scholar—Platforming the Scholarly Economy." *Internet Policy Review* 11.3 (2022).
Green, Ben, and Yiling Chen. "The Principles and Limits of Algorithm-in-the-Loop Decision Making." *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW1 Article 50 (2019): 1–24.
Hardwig, John. "The Role of Trust in Knowledge." *The Journal of Philosophy* 88.12 (1991): 693–708.
Hunt, Elle. "Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter." The Guardian, March 24, 2016, https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter.
Hussein, Eslam, Prerna Juneja, and Tanushree Mitra. "Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube." *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW1 Article 48 (2020): 1–27.
Huszár, Ferenc, et al. "Algorithmic Amplification of Politics on Twitter." *Proceedings of the National Academy of Sciences* 119.1 (2022).
Ioannidis, John P.A. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research." *Jama* 294.2 (2005): 218–228.
Introna, Lucas D., and Helen Nissenbaum. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16.3 (2000): 169–185.
Iyengar, Shanto, et al. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22 (2019): 129–146.
Jacobellis v. Ohio, 378 U.S. 184, 84 S. Ct. 1676, 12 L. Ed. 2d 793 (1964).
Jakubowicz, Andrew. "Alt_Right White Lite: Trolling, Hate Speech and Cyber Racism on Social Media." *Cosmopolitan Civil Societies: An Interdisciplinary Journal* 9.3 (2017): 41–60.
Jamieson, Kathleen Hall, and Joseph N. Cappella. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment.* Oxford University Press, 2010.
Johnson, David R., and David Post. "Law and Borders: The Rise of Law in Cyberspace." *Stanford Law Review* 48.5 (1996): 1367–1402.

Kapoor, Sayash, and Arvind Narayanan. "Leakage and the Reproducibility Crisis in Machine-Learning-Based Science." *Patterns* 4.9 (2023).
Kaptchuk, Ted J. "The Double-Blind, Randomized, Placebo-Controlled Trial: Gold Standard or Golden Calf?" *Journal of Clinical Epidemiology* 54.6 (2001): 541–549.
Keller, Daphne. "Amplification and Its Discontents: Why Regulating the Reach of Online Content is Hard." *J. Free Speech* L. 1 (2021): 227.
Kirshner, Alex. "Twitter Was for News." *Slate*, October 5, 2023, https://slate.com/technology/2023/10/elon-musk-x-twitter-news-links-headlines-why.html.
Kleinberg, Jon M. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM (JACM)* 46.5 (1999): 604–632.
Latour, Bruno, and Steve Woolgar. *Laboratory Life: the Construction of Scientific Fact*. Princeton University Press, 1986.
Laufer, Benjamin, Thomas Gilbert, and Helen Nissenbaum. "Optimization's Neglected Normative Commitments." *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (2023).
Loader, Brian D., and Dan Mercea. "Networking Democracy? Social Media Innovations and Participatory Politics." *Information, Communication & Society* 14.6 (2011): 757–769.
Le Monde. "Groupe Le Monde's Code of Ethics and Professional Conduct." 2022. https://www.lemonde.fr/en/about-us/article/2022/04/06/groupe-le-monde-s-code-of-ethics-and-professional-conduct_5979823_115.html.
Lessig, Lawrence. *Code: And Other Laws of Cyberspace*. Basic Books Inc., 1999.
Levitsky, Steven, and Daniel Ziblatt. *How Democracies Die*. Crown, 2018.
Lum, Kristian, and Lazovich, Tomo. "The Myth of the Algorithm: A System-Level View of Algorithmic Amplification." *Knight First Amendment Institute*, September 13, 2023, https://knightcolumbia.org/content/the-myth-of-the-algorithm-a-system-level-view-of-algorithmic-amplification.
Manjoo, Farhad. "Can Facebook Fix Its Own Worst Bug." *The New York Times Magazine*, April 25, 2017, https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html.
Marwick, Alice E., and Rebecca Lewis. "Media Manipulation and Disinformation Online." *Data & Society*, May 15, 2017, https://datasociety.net/library/media-manipulation-and-disinfo-online/.
Mathew, Binny, et al. "Spread of Hate Speech in Online Social Media." *Proceedings of the 10th ACM Conference on Web Science* (2019).
McCabe, David. "Lawmakers Target Big Tech 'Amplification.' What Does That Mean?" *New York Times*, December 1, 2021, https://www.nytimes.com/2021/12/01/technology/big-tech-amplification.html.
McNutt, Marcia K., et al. "Transparency in Authors' Contributions and Responsibilities to Promote Integrity in Scientific Publication." *Proceedings of the National Academy of Sciences* 115.11 (2018): 2557–2560.
Meldrum, Marcia L. "A Brief History of the Randomized Controlled Trial: From Oranges and Lemons to the Gold Standard." *Hematology/Oncology Clinics of North America* 14.4 (2000): 745–760.
Merrill, Jeremy B., and Will Oremus. "Five Points for Anger, One for a 'Like': How Facebook's Formula Fostered Rage and Misinformation." *The Washington Post*,

October 26, 2021, https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/.
Mendes, Kaitlynn, Jessica Ringrose, and Jessalynn Keller. "#MeToo and the Promise and Pitfalls of Challenging Rape Culture through Digital Feminist Activism." *European Journal of Women's Studies* 25.2 (2018): 236–246.
Mill, John Stuart. *On Liberty and Other Essays*. Oxford University Press, 1998.
Miller, Erin L. "Amplified Speech." *Cardozo L. Rev.* 43 (2021): 1.
Milli, Smitha, et al. "Twitter's Algorithm: Amplifying Anger, Animosity, and Affective Polarization." *arXiv preprint arXiv:2305.16941* (2023).
Miriam-Webster, "sophistry (n.)," https://www.merriam-webster.com/dictionary/sophistry.
Munroe, Randall. "Reddit's New Comment Sorting System." *Reddit Blog*, October 15, 2009, http://redditblog.blogspot.com/2009/10/reddits-new-comment-sorting-system.html.
Napoli, Philip, and Robyn Caplan. "Why Media Companies Insist They're Not Media Companies, Why They're Wrong, and Why It Matters." *First Monday* (2017).
Narayanan, Arvind. "Understanding Social Media Recommendation Algorithms." *Knight First Amendment Institute*, March 9, 2023, https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.
Needleman, Sarah. "Microsoft Defends New Bing after AI Chatbot Offers Unhinged Responses." *Wall Street Journal*, February 17, 2023, https://www.wsj.com/tech/ai/microsoft-defends-new-bing-says-ai-upgrade-is-work-in-progress-3447074d.
Nguyen, C. Thi. "Echo Chambers and Epistemic Bubbles." *Episteme* 17.2 (2020): 142.
Nielsen, Rasmus Kleis, and Richard Fletcher. "Democratic Creative Destruction? The Effect of a Changing Media Landscape on Democracy." *Social Media and Democracy: The State of the Field, Prospects for Reform* (2020): 139–162.
Nissenbaum, Helen. "New Research Norms for a New Medium." *The Commodification of Information* (2002): 433–457.
Noble, Safiya Umoja. *Algorithms of Oppression*. New York University Press, 2018. Benjamin, Ruha. *Race after Technology*. Polity Press, 2019.
Nosek, Brian A., et al. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115.11 (2018): 2600–2606.
Open Science Collaboration. "Estimating the Reproducibility of Psychological Science." *Science* 349.6251 (2015).
Pariser, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Books, 2012.
Phan, Thao, and Scott Wark. "Racial Formations as Data Formations." *Big Data & Society* 8.2 (2021).
Perez, Sarah. "Twitter Blocks Links to Rival Threads, While CEO Downplays Reports of Traffic Decline." *TechCrunch*, July 11, 2023, https://tcrn.ch/46GHcMP.
Phillips, Whitney. "The Oxygen of Amplification." *Data & Society*, May 22, 2018, https://datasociety.net/library/oxygen-of-amplification/.
Plato, *Sophist*, translated by Benjamin Jowett, http://classics.mit.edu/Plato/sophist.html.
Plato, *The Republic*, Book VII, https://www.gutenberg.org/cache/epub/1497/pg1497-images.html#link2H_4_0009.

Reagle, Joseph. *Good Faith Collaboration: The Culture of Wikipedia*. History and Foundations of Information Science. The MIT Press, 2010.
Riemer, Kai, and Sandra Peter. "Algorithmic Audiencing: Why We Need to Rethink Free Speech on Social Media." *Journal of Information Technology* 36.4 (2021): 409–426.
Roose, Kevin. "Don't Ban ChatGPT in Schools. Teach With It." *The New York Times*, January 12, 2023, https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html.
Roth, Emma. "Twitter Abruptly Bans All Links to Instagram, Mastodon, and Other Competitors." *The Verge*, December 18, 2022, https://www.theverge.com/2022/12/18/23515221/twitter-bans-links-instagram-mastodon-competitors.
Satariano, Adam, and Mike Isaac. "The Silent Partner Cleaning Up Facebook for $500 Million a Year." *New York Times*, August 31, 2021, https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html.
Schroeder, Milton R., and Mary M. Schroeder. "The New Encyclopaedia Britannica: All Human Knowledge." *ABAJ* 60 (1974): 711.
Shoemaker, Pamela J., and Timothy Vos. *Gatekeeping Theory*. Routledge, 2009.
Simon, Judith. "The Entanglement of Trust and Knowledge on the Web." *Ethics and Information Technology* 12.4 (2010): 343–355.
Stray, Jonathan, Ravi Iyer, and Helena Puig Larrauri. "The Algorithmic Management of Polarization and Violence on Social Media." *Knight First Amendment Institute*, August 22, 2023, https://knightcolumbia.org/content/the-algorithmic-management-of-polarization-and-violence-on-social-media.
Suarez-Lledo, Victor, and Javier Alvarez-Galvez. "Prevalence of Health Misinformation on Social Media: Systematic Review." *Journal of Medical Internet Research* 23.1 (2021).
Sunstein, Cass. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, 2017.
Tarasov, Katie. "Why Content Moderation Costs Billions and Is So Tricky for Facebook, Twitter, YouTube and Others." *CNBC*, February 27, 2021, https://www.cnbc.com/2021/02/27/content-moderation-on-social-media.html.
Taylor, C.C.W. and Mi-Kyoung Lee, "The Sophists," *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2020/entries/sophists/.
The Hindu. "Living Our Values: Code of Editorial Values." 2011. https://www.thehindu.com/news/national/living-our-values-code-of-editorial-values/article1715043.ece.
The New York Times. "Standards and Ethics." https://www.nytco.com/company/standards-ethics/.
The Wall Street Journal. "Newsroom Standards & Ethics." https://newsliteracy.wsj.com/standards-and-ethics/.
Thomas Kerchever Arnold. Spelling Turned Etymology. Gilbert & Rivington Printers, 1844.
Tucker, Joshua A., et al. "From Liberation to Turmoil: Social Media and Democracy." *J. Democracy* 28 (2017).
Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, 2017.

Van Couvering, Elizabeth. "New Media? The Political Economy of Internet Search Engines." *Annual Conference of the International Association of Media & Communications Researchers* (2004).
Vlastos, Gregory. "Socrates' Disavowal of Knowledge." *The Philosophical Quarterly* (1950–) 35.138 (1985): 1–31.
Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359.6380 (2018): 1146–1151.
Wagner, Kurt. "Facebook Says It Has Spent $13 Billion on Safety and Security Efforts since 2016." *Fortune*, September 21, 2021, https://fortune.com/2021/09/21/facebook-says-it-has-spent-13-billion-on-safety-and-security-efforts-since-2016/.
Walsch, Toby. "Gaslighting, Love Bombing and Narcissism: Why Is Microsoft's Bing AI So Unhinged?" *The Conversation*, February 17, 2023, https://theconversation.com/gaslighting-love-bombing-and-narcissism-why-is-microsofts-bing-ai-so-unhinged-200164.
Whittaker, Joe, Seán Looney, Alastair Reed, and Fabio Votta. "Recommender Systems and the Amplification of Extremist Content." *Internet Policy Review* 10.2 (2021).
Wiggers, Kyle. "Researchers Discover a Way to Make ChatGPT Consistently Toxic." *TechCrunch*, April 12, 2023, https://techcrunch.com/2023/04/12/researchers-discover-a-way-to-make-chatgpt-consistently-toxic/.


KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY