cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages X–X
https://doi.org/10.17645/mac.v11i2.6610

Commentary

The Fact of Content Moderation; Or, Let's Not Solve the Platforms' Problems for Them

Tarleton Gillespie [1,2]

[1] Microsoft Research New England, USA; tarleton@microsoft.com
[2] Department of Communication, Cornell University, USA

Submitted: 15 December 2022 | Accepted: 19 January 2023 | Published: in press

Abstract
Recent social science concerning the information technology industries has been driven by a sense of urgency around the problems social media platforms face. But it need not be our job to solve the problems these industries have created, at least not on the terms in which they offer them. When researchers are enlisted in solving the industry's problems, we tend to repeat some of the missteps common to the study of technology and society.

Keywords
content moderation; governance; industry research; platforms

Issue
This commentary is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Problem-Solving and the Fallacy of Urgency

Over the last two decades, public and scholarly attention to the information industries has expanded and deepened. Critical questions about data-centrism, privacy, inequity, labor exploitation, and new forms of monopoly power have gained traction. The area I've been most concerned with—content moderation by social media platforms—has seen an explosion of academic attention to match the explosion of concern among the public, journalists, and policymakers around the world. This is undoubtedly good news. Senior leadership at most of the major platforms seem to take trust and safety more seriously than ever before. Detection tools have become more sophisticated. More attention is being paid to the labor that moderation requires, though not enough. The individual and societal harms being perpetrated via these platforms are now understood to be urgent, in a way they were not less than a decade ago.

But this is also a public relations strategy on the part of the tech companies, a grasp for "perceived corporate authenticity" (Hanlon & Fleming, 2009). Whenever troubled industries begin to acknowledge the concerns of their customers, they step up their corporate social responsibility efforts, recommit to the health of the public, the environment, the labor force, etc., and downplay the tactical value of these gestures. Even if well-intentioned, these gestures help to stabilize the understanding of the problems at hand, valorize the role of those companies in addressing them, demarcate appropriate solutions, and normalize the relations between the company, public, market, and state on which they depend (Baker & Hanna, 2022; Busch & Shepherd, 2014). The professionalization of Trust & Safety inside the companies for example, which is a welcome change in general, has also affirmed specific approaches to content moderation—reifying who counts as users, what registers as legitimate harm, and what reads as a reasonable intervention. And, as has been most painfully apparent in Elon Musk's takeover of Twitter in 2022, the content moderation debate can itself be deployed as a political cudgel, turning the problem into a game of "should he or shouldn't he."

These efforts also reshape what kinds of research get done, and what kinds enjoy the greatest visibility. As public and regulatory scrutiny has intensified, it is

the administrative, problem-solving research that has increasingly taken center stage: legal and economic analysis displace the sociological; American cases displace comparative ones; data-centric efforts to measure problems displace interpretive efforts to investigate them. Industry-friendly analysis crowds out the critical, the feminist, and the postcolonial, anything that might challenge the industry itself.

The widely shared sense of urgency about these issues—an urgency academics feel too, as users and citizens—has also drawn some researchers into lockstep with the social media companies, privileging a problem-solving mentality that takes for granted the definition of the problem and the aims of the stakeholders. Junior scholars are being lured by funded research projects and cajoled into taking unpaid advisory roles by the platform teams facing these controversies. Funding organizations have poured money into what Anderson (2021, p. 44) calls "consequence-driven and interpretation free" research on digital media and its effects. Funders like to see direct "engagement" with the companies as evidence of impact. Industry-academic partnerships, journals, and conferences enjoy outsized prominence, focusing attention on measuring and reducing harm while overshadowing equally important research about labor and inequity, subcultural expression, and the alternative approaches to moderation being squeezed out of view.

So social scientists, take note: It need not be our job to solve the problems these industries have created, at least not on the terms in which they offer them. Given my own employment, I get how odd, maybe problematic, it is for me to say this (see, for example, Sætra et al., 2022). But research oriented to problem-solving, while it may solve problems, also accepts the questions posed by industry stakeholders themselves—questions that are by no means innocent.

This may seem to run counter to the widespread concern that social media companies have frustrated academics by withholding access, to the massive data and rarified computational systems needed to conduct their research (Couch, 2020; Social Science Research Council, 2018), and with NDA arrangements deployed so as to thwart qualitative inquiry (Starr, 2020). But I believe the concern is in fact the same. Access to data is just not a wall, but a gate. By doling out access churlishly, the tech industry often can "capture" those researchers they do interact with, drawing a select few academic institutions into a cozy orbit (Whittaker, 2021)—a coziness that can leave people suspicious of the research they subsequently generate (Matias, 2020). And access to the people and inner workings of these companies can also be granted and withheld in ways designed to protect them from unfavorable assessment by researchers.

So I don't begrudge the field's demands for access to social media data. And I share the growing concerns about research partnerships with Silicon Valley. Mine is a complementary concern: Even researchers who do not enjoy access can nevertheless be captured—by embracing a problem-solving orientation that accepts the way the tech industries define their problems. This is akin to what Gitlin (1978)—commenting on mass media research of the time—called an "administrative mentality": research that "poses questions from the vantage of the command-posts of institutions that seek to improve or rationalize their control over social sectors in social functions" (1978, p. 225). Silicon Valley companies need researchers now more than ever, as signals of their good faith efforts, as they face multiple crises that have stirred public discontent and regulatory scrutiny that threaten their very existence. Plus, their ability to enlist researchers is stronger than ever, as they play on our genuine concern for the public welfare—that these companies jeopardized (Benson & Kirsch, 2010).

2. The Fact that Content Moderation Exists

Solving the problems the industry created on the terms they offer can lead us to overlook the problems we are not being invited to solve, the communities the industry tends to ignore, the solutions that challenge the business models embraced by the industry, and those dilemmas that are in fact not solvable, but are actually meant to be perennially contested. We are kept from thinking about how else moderation might be, or how the very fact of content moderation configures public power.

This kind of work is being done, certainly. Over more than a decade, scholars have rightly focused on expanding our understanding of the practices and dynamics involved in content moderation. Researchers in information studies, communication and media, and sociology have considered the entire sociotechnical ensemble being fixed into place: technological, institutional, social, and legal.

But when we adopt a problem-solving approach on terms borrowed from social media companies, we risk accepting as a precept that content moderation exists, and must exist in the way that it does—to accept that social media exists in the way it does. Even as this field grows more impactful, it takes too much for granted. This is an enormous mistake. Because it may be that what will matter about social media platforms and other information industries—their lasting significance—will not be about the specific dilemmas the technology sparks, or how well or poorly the industry stakeholders address them. It may be the very fact of content moderation as a societal project, the very fact of these industries and the roles they have inhabited, to which we should attend. "The fact of" is borrowed from Cavell's 1982 essay "The Fact of Television," and even more so from Streeter's (1996) use of Cavell in the start of his book *Selling The Air*.

Content moderation is an illustrative example. We can debate the facts *about* content moderation, how content moderation is or should be done, and how harm is or should be addressed. A judge in Texas may want there to be less moderation, while a feminist activist

harassed by misogynist trolls may want more. But both positions require there to be moderation, of some sort, and that means some apparatus that can accomplish moderation. Do the kinds of decisions platforms make matter? Of course. Does the scope of the specific problems they face matter? Absolutely. But it also matters that Silicon Valley has assembled an enormous labor force to do the work of moderation that didn't exist before, fitted with specific labor dynamics. It matters that the imposition of content moderation is driving some users to alt-sites that assert different moderation policies, cleaving political discourse in a particular way. It matters that the idea of moderation has enhanced and altered the cultural power of Silicon Valley companies. It matters that long-standing theories of regulation are shifting when it comes to the technology industry. It matters that moderation is helping reassert a new form of American cultural imperialism, under the guise of care. It matters that debates about content moderation and the responsibility of platforms have forced a subtle redefinition of "media" itself. Only research scoped so as to take in that entire sociotechnical ensemble shifts our attention from how moderation is done, to the very fact *of* it.

Studying the fact of content moderation means paying attention to what would still matter even if all the social media platforms disappeared. New legal and regulatory regimes are not only imposing obligations on social media platforms, but they're also generating new government agencies, new policy techniques, and whole new categories into which information intermediaries must adjust to fit. Tech firms are generating new managerial positions and adjudicative practices and forms of expertise inside the companies. Some are even attempting to manufacture institutional forms where none existed: Facebook's Oversight Board, which is designed to appear like an institutional partnership, though it is more like Facebook extruding out a part of itself so as to partner with it, may end up a model for other companies.

Or consider everyone being enlisted by the platform companies into playing sustained roles in the project of content moderation: non-profits and advocacy groups, content creator coalitions, institutional partners like fact-checking teams from news organizations, law enforcement, and regulatory agencies (Ananny, 2018). What is the relationship between these organizations and the platforms? Who defines and funds their efforts? What financial and political pressure does this impose on them, and how do they bear that pressure? What legitimacies and expertise are being called for and brought to bear, and how is all that borrowed authority used to legitimate the entire undertaking? How does their partnership with platforms, such as it is, alter how they understand their own public mission?

What social media is remains unsettled. What will later seem true about it may have more to do with the shifting institutions and arrangements being pulled together to stabilize it. The consequences that we will later mistakenly attribute to the technology of social media platforms will depend on the growth of new institutions, and the adjustments of existing institutions, around platforms and their governance (Johns, 1998). These arrangements will almost surely outlast the platform companies themselves.

3. We Are Implicated

When we focus on solving platforms' problems for them, our critical attention is obscured: by the sense of urgency, by the tactical way platforms define the problems for us, and by the fetishization of data science approaches and scalable solutions. Our concerns are replaced by theirs, or by none at all: Too often, calls for social science to be more oriented toward problem-solving seem to demonstrate a troubling disregard as to where those problems come from (Watts, 2017). I am not saying that the critical study of a datafied society needs to be irrelevant or utopian. But as researchers, we can and should opt to stand aside from this rush to solve immediate problems, to instead ask questions about the underlying dynamics that manifest themselves in these problems, about what the problem assumes or overlooks—and about what arrangements might provide solutions far down the road, but have to wait until our current institutional commitments have shifted (Splichal, 2008). While today's problems are urgent, they've been urgent far longer than the social media companies that now face them have been around.

In fact, if we take seriously the idea that it is the arrangements of institutions that will matter in the long run, then being enlisted in solving platforms' problems is us becoming part of that arrangement. We are implicated because we are among those being ensnared in this institutional and sociotechnical ensemble. If "don't solve problems" sounds counterintuitive, it is a reminder of how much our fields have already been enlisted in this project, by Silicon Valley and by the public outcry. Whatever the future fortunes of Facebook, or Meta, or whatever, the proximity between researchers and social media companies itself will matter: the implicit agreements being established about who takes on what responsibilities, who bears what costs, and who defines which goals.

Now, if we do refuse to dutifully engage in agnostic problem-solving, we still have a duty to meet with our research:

> Critical theory is, of course, not unconcerned with the problems of the real world. Its aims are just as practical as those of problem-solving theory, but it approaches practice from a perspective which transcends that of the existing order, which problem-solving theory takes as its starting point. (Cox, 1981, p. 130)

We should aspire to offer insights into the deeper assumptions embedded in how the problems are defined, the very fact of these sociotechnical systems as a part of the world, and the institutional arrangements being fixed into place. Understanding these, in their historical, sociological, and political-economic contexts, can be put into the service of more profound changes.

Acknowledgments

My gratitude to Elizabeth Fetterolf for their superb research support, and to Robyn Caplan, Mary Gray, and Dylan Mulvin for their comments on drafts of this essay.

Conflict of Interests

It is worth reiterating that I am employed by Microsoft, in the Microsoft Research lab in New England. The opinions expressed here are my own.

References

Ananny, M. (2018). *The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation*. Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/D85B1JG9

Anderson, C. W. (2021). Fake news is not a virus: On platforms and their effects. *Communication Theory*, *31*(1), 42–61. https://doi.org/10.1093/ct/qtaa008

Baker, D., & Hanna, A. (2022, June 7). AI ethics are in danger. Funding independent research could help. *Stanford Social Innovation Review*. https://doi.org/10.48558/VCAT-NN16

Benson, P., & Kirsch, S. (2010). Capitalism and the politics of resignation. *Current Anthropology*, *51*(4), 459–486. https://doi.org/10.1086/653091

Busch, T., & Shepherd, T. (2014). Doing well by doing good? Normative tensions underlying Twitter's corporate social responsibility ethos. *Convergence*, *20*(3), 293–315. https://doi.org/10.1177/1354856514531533

Cavell, S. (1982). The fact of television. *Daedalus*, *111*(4), 75–96.

Couch, C. (2020). *The data driving democracy: Understanding how the internet is transforming politics and civic engagement*. American Academy of Arts & Sciences.

Cox, R. W. (1981). Social forces, states and world orders: Beyond international relations theory. *Millennium: Journal of International Studies*, *10*(2), 126–155. https://doi.org/10.1177/03058298810100020501

Gitlin, T. (1978). Media sociology: The dominant paradigm. *Theory and Society*, *6*(2), 205–253.

Hanlon, G., & Fleming, P. P. (2009). Updating the critical perspective on corporate social responsibility. *Sociology Compass*, *3*(6), 937–948. https://doi.org/10.1111/j.1751-9020.2009.00250.x

Johns, A. (1998). *The nature of the book: Print and knowledge in the making*. University of Chicago Press.

Matias, J. N. (2020). *Why we need industry-independent research on tech & society*. Citizens and Technology Lab. https://citizensandtech.org/2020/01/industry-independent-research

Sætra, H. S., Coeckelbergh, M., & Danaher, J. (2022). The AI ethicist's dilemma: Fighting big tech by supporting big tech. *AI and Ethics*, *2*(1), 15–27. https://doi.org/10.1007/s43681-021-00123-7

Social Science Research Council. (2018). *To secure knowledge: Social science partnerships for the common good*. https://www.ssrc.org/to-secure-knowledge

Splichal, S. (2008). Why be critical? *Communication, Culture & Critique*, *1*(1), 20–30. https://doi.org/10.1111/j.1753-9137.2007.00003.x

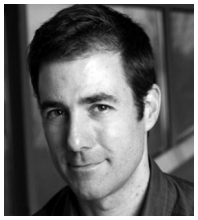
Starr, P. (2020, January 22). How money now tries to bury the truth. *The American Prospect*. https://prospect.org/power/how-money-now-tries-to-bury-the-truth

Streeter, T. (1996). *Selling the air: A critique of the policy of commercial broadcasting in the United States*. University of Chicago Press.

Watts, D. J. (2017). Should social science be more solution-oriented? *Nature Human Behaviour*, *1*(1), Article 0015. https://doi.org/10.1038/s41562-016-0015

Whittaker, M. (2021). The steep cost of capture. *Interactions*, *28*(6), 50–55. https://doi.org/10.1145/3488666

About the Author



Tarleton Gillespie is a senior principal researcher at Microsoft Research, and an affiliated associate professor in the Department of Communication and Department of Information Science at Cornell University. He is the author of *Wired Shut: Copyright and the Shape of Digital Culture* (MIT, 2007), co-editor of *Media Technologies: Essays on Communication, Materiality, and Society* (MIT, 2014), and author of *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* (Yale, 2018).