

Internal Fractures: The Competing Logics of Social Media Platforms

Angèle Christin¹ , Michael S. Bernstein¹, Jeffrey T. Hancock¹,
Chenyan Jia², Marijn N. Mado¹ , Jeanne L. Tsai¹, and Chunchen Xu¹

Social Media + Society
July-September 2024: 1–14
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20563051241274668
journals.sagepub.com/home/sms



Abstract

Social media platforms are too often understood as monoliths with clear priorities. Instead, we analyze them as complex organizations torn between starkly different justifications of their missions. Focusing on the case of Meta, we inductively analyze the company's public materials and identify three evaluative logics that shape the platform's decisions: an *engagement* logic, a *public debate* logic, and a *wellbeing* logic. There are clear trade-offs between these logics, which often result in internal conflicts between teams and departments in charge of these different priorities. We examine recent examples showing how Meta rotates between logics in its decision-making, though the goal of engagement dominates in internal negotiations. We outline how this framework can be applied to other social media platforms such as TikTok, Reddit, and X. We discuss the ramifications of our findings for the study of online harms, exclusion, and extraction.

Keywords

social media platforms, evaluation, competing logics, engagement, public debate, wellbeing

Introduction

As social media companies face increasing public scrutiny, it has become difficult to pinpoint how, exactly, they justify their priorities. Take the case of Meta and Facebook: “Facebook was built to bring people closer together,” proclaims a blog post from the previous head of Facebook's Newsfeed (Mosseri, 2018). “Meta Is Advancing Democracy,” announces another company blog post, despite evidence that its first goal—bringing friends closer together—can have harmful effects on democratic discourse (Potts, 2023). “Making Emotional Health a Priority,” explains a third public post (Meta, 2021b), against a backdrop of concerns about bullying, peer comparison, as well as racist and sexist harassment (Wells et al., 2021).

This revolving set of justifications reflects the muddled infrastructures that platforms create. As socio-technical systems, social media companies are subject to shifts in norms and decisions as developers, users, and policies co-evolve (Gillespie, 2018). Yet even as social media companies tweak their algorithms, change their content moderation policies, develop fact-checking partnerships, send mental-health-related notifications, and create transparency centers, their responses are criticized as piecemeal, incomplete, and inefficient (Ananny, 2018; Hao, 2021; Marwick, 2021; Napoli & Caplan, 2017; Sharp & Gerrard, 2022; Sobieraj, 2020). In

this article, we suggest that these priority paradoxes and incomplete implementations come into focus if we analyze the struggles and conflicts that characterize how social media companies justify their mission. We understand social media companies as complex organizational entities with internal tensions, different priorities, and distinct teams and departments in charge of these goals.

Specifically, we ask: how do different systems of values *compete and conflict* at social media companies? We focus on Meta's (Facebook, Instagram, Threads, and WhatsApp) public material and inductively identify three distinct ways in which Meta justifies its policies. First, we highlight the dominance of an *engagement logic* based on the optimization of users' signals, which maximizes online advertising revenues and the platform's broader commercial goals. Second, we identify a *public debate logic*, with an emphasis on verified information, rational-critical discourse, and pro-democratic outcomes. Third, we delineate a *wellbeing logic*

¹Stanford University, USA

²Northeastern University, USA

Corresponding Author:

Angèle Christin, Department of Communication, Stanford University, Room 334, Building 120, 450 Jane Stanford Way, Stanford, CA 94305, USA.

Email: angelec@stanford.edu



centered on promoting and repairing users' psychological and physical health. We then outline the applicability of this framework to other social media platforms.

These different justifications are often in direct conflict with one another in the daily internal operations of social media companies. Drawing on the documents made public by Frances Haugen in the 2021 Facebook leaks, we find that public debate and wellbeing typically remain subordinate to engagement during key decision points. We argue that the visible presence of the public debate and wellbeing logics is what enables social media top executives to publicly save face and continue doing "business as usual," even when their products are consistently criticized because they amplify racialized, gendered, and other forms of injustice. We conclude by discussing the ramifications of these findings for the study of online inequality and extraction.

From Values to Competing Logics: Making Sense of Social Media Platforms

Broadly considered, social media platforms are a set of websites and applications characterized by users' ability to create individual accounts, post content on their accounts, and share content with lists of contacts, both public and private (boyd & Ellison, 2007). While most social media companies have long described themselves as neutral actors, scholars have highlighted the political, economic, and cultural values embedded in their design and policies. Overall, however, social media companies are too often perceived as monoliths without internal complexity. Drawing on organizational sociology, we outline the benefits of examining the role of internal conflicts, fractures, and tensions in how social media companies justify their priorities.

Platforms and Their Values

Over the past 20 years, social media platforms have repeatedly sought to position themselves as neutral intermediaries. As Gillespie (2010) first pointed out, the metaphor of the platform itself downplays these companies' role, presenting it as "just hosting" digital content and conveniently suggesting a "hands-off neutrality" (p. 358). Social media platforms insisted early on that they were not media companies, largely because media companies are obliged to police various types of speech and have historically experienced more intensive government oversight (Helft, 2008; Napoli & Caplan, 2017). To date, the U.S. judicial system has upheld this framing: Instead of grouping platforms with newspapers or broadcasters, legislation has treated platforms as information conduits comparable to telecom companies and post offices. Following Section 230 of the 1996 Telecommunications Act, social media platforms are protected from legal liability for any hosted, curated, or distributed third-party speech.

Against this statutory neutrality, scholars have pointed out that companies such as Meta and Google are in fact media companies (Napoli & Caplan, 2017). Platforms constantly tinker with the flow of online information: They determine "how profiles and interactions are structured; how social exchanges are preserved; how access is priced or paid for; and how information is organized" (Gillespie, 2018, p. 22). For instance, when Instagram allowed users to edit photos through filters, this directly shaped content-production strategies (Leaver et al., 2020). Humans and algorithms also fulfill editorial functions on social media platforms through content moderation (Klonick, 2018; Roberts, 2021).

What are the values underlying the interventions and policies of social media platforms? Scholars have examined this question by exploring how platform design and content moderation shape the experience of marginalized users (Benjamin, 2019; Brock, 2020; Bucher, 2021; Chander & Krishnamurthy, 2018; Noble, 2018; van Dijck et al., 2018; van Dijck & Poell, 2013). Brock (2020) notes how the alleged neutrality of platforms relies on the enactment of "color-blind" norms that position Whiteness as the default internet identity, while Noble (2018) highlights the racist and sexist bias of search engine algorithms. Scholars also find that platforms' commitment to neutrality often amplifies the spread of misogynistic online content. For Lewis (2018), platforms' "attempts at objectivity are being exploited by users" (p. 44). Massanari (2017) adds: "remaining 'neutral' in these cases valorizes the rights of the majority while often trampling over the rights of others" (p. 339). These studies remind us that when platforms seek to remain "neutral" and "objective" in contested political spheres, they are in fact making political choices.

To explain these choices and the values of social media platforms, scholars have mobilized two complementary arguments. First, they emphasize the lack of diversity within the platforms' labor force. As Gillespie (2018) writes, "the full-time employees of most social media platforms are overwhelmingly white, overwhelmingly male, overwhelmingly educated, overwhelmingly liberal or libertarian, and overwhelmingly technological in skill and worldview" (p. 12). This sociodemographic base is often invoked to explain why platforms' executives and engineers keep searching for "neutrality" and "objectivity," even when the harms associated with adopting such a "view from above" or "god trick" are well documented (Haraway, 1988; Hill Collins, 2000). The background and values of executives and engineers, together with the project-based internal structure of most large technology companies, further thwart the efforts of "ethical entrepreneurs"—often women and people of color—seeking to change the status quo (Ali et al., 2023).

Second, scholars highlight the economic incentives of platforms to understand their policy and design choices. For Couldry and Mejias (2019a, 2019b), social media platforms perform "data colonialism" through their collection, mining, and selling of user data. In this view, Facebook is a "principal actor in data colonialism" because it is one of the

“corporations involved in capturing everyday social acts and translating them into quantifiable data which is analyzed and used for the generation of profit” (Couldry & Mejias, 2019b, p. 340). Thus, social behaviors and interactions that previously existed outside of marketplaces are now perceived as “raw materials” appropriated to extract and create revenue by selling behavioral data to advertisers (Zuboff, 2019). This data mining business model prompts platforms to boost user engagement. As a result, social media companies actively draw users to the platform and keep them there through personalized algorithmic targeting.¹

From rewarding sexist content on algorithmic feeds to monetizing user data through behavioral targeting, social media platforms are always in the process of shaping user interactions in ways that reveal their underlying economic, political, and social values. To date, scholarship examining these values has primarily focused either on the macro, societal level (showing how social media platforms encode broad racist, misogynistic, and capitalistic values in their design) or at the micro, individual level (examining the negative individual experiences of marginalized users on social media). To explain these values, scholars have turned to two complementary arguments, one at the micro level (technology companies’ lack of diversity) and one at the macro level (the economic incentives of platforms).

Competing Logics

Our intervention is to look at the meso, organizational level and move beyond unitary analyses of social media platforms as monoliths driven by a single mission. Instead, we understand social media companies as complex entities with internal tensions and fractures; different priorities represented by various employees, projects, and departments; and distinct ways of justifying these goals. Specifically, we ask: how do different systems of values *compete and conflict* in how social media companies justify their priorities?

To examine these tensions, we turn to organizational sociology and the sociology of evaluation (Lamont, 2012). In organizational sociology, multiple studies have analyzed how different norms, rules, and values compete within existing fields and organizations. Scholars developed the concept of “institutional logics” to describe these broad systems of values (Friedland & Alford, 1991). Thornton and colleagues (2012) define institutional logics as the “socially constructed, historical patterns of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organize time and space, and provide meaning to their social reality” (p. 51). The conflict and replacement between logics has been studied across sites. For instance, Thornton and Ocasio (1999) document how U.S. academic publishing switched in the 1970s from a primarily editorial orientation focused on prestige and personal imprints to a market logic focused on high profit margins and market share.

Yet institutional logics also have been criticized as deterministic and unitary. As Fligstein and McAdam (2012) noted, “The use of the term ‘institutional logics’ tends to imply way too much consensus in the field about what is going on and why and way too little concern over actors’ positions” (p. 11). A distinct approach emerged with the sociology of evaluation, which examines “how an entity attains a certain type of worth” (Lamont, 2012, p. 205). In this view, evaluative practices are always deeply contested processes. For instance, Boltanski and Thévenot (2006) delineate six broad “orders of worth”—*inspired, domestic, civic, opinion, market, and industrial*—that people use when they try to justify their understanding of a situation. They explicitly focus on the coexistence of different orders and the conflicts that emerge between them; they study controversies where people struggle over different understandings of the same context.

Drawing on the sociology of evaluation, scholars have examined the negotiations and evaluative practices through which people and groups struggle to define “what counts” (Beckert & Aspers, 2011; McPherson & Sauder, 2013). For instance, Stark (2009) offered the concept of “heterarchy” to describe organizations where different orders of worth coexist and where accountability is distributed. Similarly, Christin (2020) argues that workers and organizations are torn between distinct “modes of evaluation,” which she defines as “all the cognitive, discursive, and practical operations by which people categorize and hierarchize ideas, objects, and practices” (Christin, 2020, p. 71). In the web newsrooms she studied, journalists moved back and forth between editorial and click-based definitions of their work. Here, we build on these studies that seek to understand the internal contradictions, struggles, and negotiations shaping how organizations justify their mission to analyze social media companies.

Methods and Data

In this article, we focus on the “discursive work of justification” (Gibson et al., 2023) of social media platforms, with a specific emphasis on the public material produced by Meta, and to a lesser extent by TikTok, Reddit, and X (formerly known as Twitter).

Meta (previously Facebook) is a U.S. technology conglomerate that owns Facebook, Instagram, Threads, and WhatsApp, among other services. Facebook launched in 2004 and became accessible to anyone with an email address in 2006. The company filed for an Initial Public Offering (IPO) in 2012 and acquired Instagram the same year. As of 2024, Facebook is the most used social media platform worldwide, with 3 billion monthly active users (Dixon, 2024).

From Edward Snowden’s 2013 revelations about the Prism program to backlash against the “emotional contagion” experiment in 2014, the contested role of Facebook during the 2016 presidential campaign, the Cambridge Analytica scandal in 2018, CEO Mark Zuckerberg’s testimony to

Table 1. The Logics of Engagement, Public Debate, and Wellbeing.

	Engagement	Public debate	User Wellbeing
Goals	Maximize user engagement (DAU, MSI, etc.) Online advertising revenue (based on engagement and reach)	Foster political debate based on accurate information Protect election integrity A digital “public square”	Decrease “toxicity” online Maximize users’ subjective wellbeing as well as mental and physical health
Features and policies	Recommendation algorithms “Infinite scroll” Constant notifications “Growth hacking” Behavioral tracking	Fact-checking partnerships Community guidelines ban dangerous organizations and misinformation Interventions to maximize voter turnout “Social feedback” ordering	Content moderation guidelines ban harassment and self-harm Politics and news not promoted Hiding the “like” count
Perception of users	Sources of data Advertising audience	Information consumers Active citizens and voters Deliberative public	Individuals with psychological and physical needs “Meaningful” connections in “communities”
Core metrics	Engagement metrics Revenues, stock prices	Voter turnout Quality scores	Experiments Satisfaction surveys
Broader logic and references	Market logic (Boltanski & Thévenot, 2006), data colonialism (Couldry & Mejias, 2019a, b).	Civic logic (Boltanski & Thévenot, 2006), democratic public sphere (Habermas, 1989)	Domestic logic (Boltanski & Thévenot, 2006), logic of care (Mol, 2008; Tronto, 2013)

Congress in 2018 (and again in 2024), and the revelations of whistleblower Frances Haugen about internal dynamics at Facebook and Instagram in 2021, Meta has been at the center of many controversies (Silverman, 2016; Wells et al., 2021). The company responded to these scandals by adjusting its algorithms and content-moderation guidelines. In the process, the top executives often explained and justified their decisions, providing valuable public material within a broader context of corporate secrecy.

Meta is undoubtedly the most studied social media platform (Bucher, 2021; Frenkel & Kang, 2021; Frier, 2020; Vaidhyanathan, 2018), probably due to its age, size, and controversial role. This sizable coverage, together with the controversies surrounding Meta, makes it a relevant case to see how the evaluative logics framework helps to make sense of the company’s justifications over time. Thus, we analyzed all of Meta’s publicly available materials from 2010 to present about its policies, algorithms, experiments, and content moderation guidelines, together with the existing literature on the platform. We combed through Meta’s blog posts, high-profile as well as less-visible media interviews with employees, the company’s 2012 IPO Securities and Exchange Commission (SEC) registration statement, technical research papers published by scientists working for Meta or with Meta’s data, internal research leaked to the media, and secondary analyses.

Building on Bucher’s analysis of Facebook’s “tensions and transitions” (Bucher, 2021, pp. 2–32) and related work on the multiple metaphors inherent to platforms’ slippery public presentation (Gibson et al., 2023; Hoffmann et al., 2018; Seaver, 2017), we paid close attention to the contrasts, changes, and switches in the discourses and justifications offered by Meta’s executives and employees. We also

examined the negotiations and tensions that publicly emerged between different teams, departments, and top executives.

Despite its size and impact, Meta is only one of the many social media companies operating in contemporary societies—and it may well be idiosyncratic (Bucher, 2021). Thus, we turned to other social media platforms to find out if similar repertoires and tensions emerged. We reviewed the public material and executive statements provided by TikTok, Reddit, and X (formerly known as Twitter), repeating the process we had developed to analyze Meta.

Mapping the Competing Logics of Social Media Platforms: Meta and Beyond

Based on this two-staged examination, we identified three public definitions of social media companies’ mission: a definition based on *engagement*, a definition based on *public debate*, and a definition based on *user wellbeing* (see Table 1). In the rest of this article, we call these distinct priorities “logics” that offer somewhat consistent blueprints for evaluation, action, and justification. Yet we also acknowledge that the use of the term “logic” tends to reify dynamics that are always contentious, multiple, and temporary. Our overview of these logics is ideal-typical, in the sense that these are simplified constructs offered for the purpose of analytical clarity; the reality of their implementation is messier. We first outline what these distinct logics look like in the case of Meta, before briefly turning to TikTok, Reddit, and X.

Engagement

The first priority emerging from Meta’s public material is all about user engagement. When presenting itself to potential

advertisers, the company's commercial page states: "Your customers are here. Find them with Meta Ads." The page further offers to "drive engagement" and "optimize for link clicks" by inviting brands to "show ads to people likely to be interested in your business and get more messages, video views or post engagement" (Meta, 2024).

This rhetoric of engagement is Meta's financial backbone. As early as 2012, in the company's IPO SEC filing document, an entire section entitled "How We Create Value for Advertisers and Marketers" (p. 75) emphasized the centrality of user engagement as a key value proposition (Ebersman, 2012). After noting that "Facebook offers the ability to reach a vast consumer audience of over 800 million Monthly Active Users (MAUs) with a single advertising purchase" (p. 75), the report turned to the unique strength of the platform, namely its system of "social ads," where brands could opt to display their ads with "social context" (additional text showing which friends have "liked" or otherwise interacted with the brand). Facebook's SEC statement repeatedly highlighted the role of engagement among "fans" (a term frequently used in the statement) in terms of advertising reach:

We believe that the shift to a more social web creates new opportunities for businesses to *engage* with interested customers. Most of our ad products offer new and innovative ways for our advertisers to interact with our users, such as ads that include polls, encourage comments, and invite users to an event [. . .]. We believe that Page owners can use Facebook ads and sponsored stories to increase awareness of and *engagement* with their Pages. (Ebersman, 2012, p. 77, emphasis added)

Over the years, Meta developed a range of metrics to measure user engagement. These include the number of MAUs, Daily Active Users (DAU), Meaningful Social Interactions (MSI), as well as more granular metrics. To optimize user engagement metrics, Facebook has been implementing a range of technical features. A significant innovation in this logic is the adoption of the "infinite scroll," a 2006 tweak in which the timeline never ends and users can keep on browsing infinitely. Aza Raskin, a Silicon Valley designer widely credited with the invention of the infinite scroll, explained in 2018 (Andersson, 2018; see also Cohen, 2021):

It's as if they [Facebook and other social media platforms] are taking behavioral cocaine and just sprinkling it all over your interface and that's the thing that keeps you like coming back and back and back [. . .]. Behind every screen on your phone, there are generally like literally a thousand engineers that have worked on this thing to try to make it maximally addicting [. . .]. If you don't give your brain time to catch up with your impulses, you just keep scrolling.

Meta also adopted other strategies to increase engagement. For instance, since its early days, Facebook has been hiring "growth hackers": employees tasked with designing and

testing approaches to build a larger user-base, ultimately growing overall engagement (Fidelman, 2013). The company also developed a range of technical nudges, including recommendation algorithms that primarily draw on users' previous engagement signals (Meta, 2019); notifications to bring users back to the platform when there is activity on their posts (Meta, 2023); and the widespread deployment of "likes" despite internal research that found such "likes" made it hard for users to log off.

In official comments, Facebook denied ever seeking to make the platform addictive to users. In 2018, a representative stated that the platform was designed to "bring people closer to their friends, family, and the things they care about [. . .]. At no stage does wanting something to be addictive factor into that process" (Andersson, 2018). Yet former employees regularly contradicted this statement. For example, in 2017, former Facebook president Sean Parker stated that Facebook would send "a little dopamine hit every once in a while [. . .] in the forms of 'likes' and comments. The goal was to keep users glued to the hive, chasing those hits while leaving a stream of raw materials in their wake" (cited in Zuboff, 2019, p. 451). Quotes by executives further reveal a wide acceptance of the priority of engagement at Meta (Rhodes & Orłowski, 2020).

This logic of engagement is obviously predicated on Meta's reliance on online advertising, but also on a broader quest to maximize shareholder and investor value. Not only do platforms have to retain engaged users for advertising purposes, they also need growing metrics to attract funding. As Raskin, the infinite scroll designer, put it (Andersson, 2018): "To get the next round of funding, to get your stock price up, the amount of time that people spend on your app has to go up [. . .]. So, when you put that much pressure on that one number, you're going to start trying to invent new ways of getting people to stay hooked."

Returning to the sociology of evaluation, this focus on engagement echoes what Boltanski and Thévenot define as a broader *market* logic, organized around values of investor capitalism, efficient transactions, and share prices (Boltanski & Thévenot, 2006; Thornton et al., 2012). But it also differs from this traditional commercial ideology in its emphasis on user participation, metrics, and data, which resonates with Couldry and Mejias' (2019a, 2019b) analysis of "data colonialism": This monetization of "the social" is the business model that social media platforms have successfully enabled and commercialized over the past 20 years, with Meta leading the pack (see also Turow, 2011; Zuboff, 2019).

Perhaps unsurprisingly, this engagement-driven language is not highly publicized on Meta's front webpages: It is primarily directed at potential investors, brands, and customers or criticized by former employees. The next two logics, in contrast, are at the front and center of Meta's public presentation of itself.

Public Debate

The second priority found in Meta's public material centers on public debate—how to improve its quality online and its connection to the democratic process. In Facebook's early years, Zuckerberg frequently mentioned the importance of “free expression” on and through the platform (Bucher, 2021, p. 162). In recent years, executives have switched the emphasis to the term “public debate”—perhaps to avoid some of the connotations of free speech absolutism. For instance, in a blog post, Meta's President of Global Affairs Nick Clegg wrote about the company's responsibility to protect “public debate” and explicitly linked the exchanges taking place on the platforms (“messy” but “positive”) to people's democratic attitudes (Clegg, 2023).

Across Meta's public material, three points emerge. First, Meta repeatedly anchored its discussion of online public debate to the question of misinformation and disinformation. Following the 2016 U.S. election cycle (Silverman, 2016), Meta ramped up its partnerships with non-partisan organizations such as Snopes and the International Fact-Checking Network to identify misinformation on Facebook and Instagram (Meta, 2022a, 2023b). Meta also sought to ban other types of problematic content, including violent content and hate speech; inauthentic content (manipulated by political actors); content produced by state-controlled media in non-democratic countries; and content interfering with democratic election processes (Facebook, 2023).

Second, Meta embraced a paternalistic responsibility to “protect” the integrity of the democratic electoral process, establishing a direct connection between information seeking on their platforms and people's political behavior. Many programs were developed for the 2020 U.S. Presidential Election. For instance, in 2019, Meta executives issued the following statement: “We have a responsibility to stop abuse and election interference on our platform. That's why we've made significant investments since 2016 to better identify new threats, close vulnerabilities and reduce the spread of viral misinformation and fake accounts” (Rosen, 2019). Meta also developed interventions to nudge users to vote through its “Voting Information Center” on Facebook and Instagram, which provided detailed information about how to vote and showed posts from verified election officials. Banners displayed countdowns to election day, options to share the countdown with friends, reminders about voting procedures, and calls to sign up as a poll worker in one's state (Gleit et al., 2020). In an op-ed, Zuckerberg (2020) reiterated the responsibilities of the company:

People want accountability, and in a democracy, the ultimate way we do that is through voting. With so much of our discourse taking place online, I believe platforms like Facebook can play a positive role in this election by helping Americans use their voice where it matters most—by voting. We're announcing on Wednesday the largest voting information campaign in American

history. Our goal is to help 4 million people register to vote. As we take on this effort, I want to outline our civic responsibilities.

Of course, Zuckerberg's statement about the “positive role” of Facebook on democracy should be taken with a grain of salt. For instance, many experts have criticized the efficacy of Meta's fact-checking initiatives (Ananny, 2018) and the problematic impact of Meta's policies in non-democratic countries (Amnesty International, 2022).

More broadly, Meta tackled the question of the “quality” of public debate on its platforms. This comes through in several publications about “discussion quality” in the “digital public square” (Facebook, 2017). For instance, a 2017 study using Meta data analyzed how the ordering of comments on a Facebook post may affect the “quality” of comments—defined as “in-depth, interesting, engaging statement or question that is worth reading, and adds to the comment conversation in an exceptional or noteworthy way” (Berry & Taylor, 2017, p. 1373).

Across these interventions, online users are not described in terms of their spending power or time engaged on the platform, but rather as citizens and voters who seek information online and whose exchanges, attitudes, and behaviors need to be protected for the democratic polity to flourish. Interestingly, on paper at least, these statements seem to be in line with the role that van Dijck et al. (2018) envision for Western European governments, rather than support the image of platforms as neoliberal or libertarian organizations (Couldry & Mejias, 2019a, 2019b; Zuboff, 2019). Meta executives emphasize their state-like responsibilities, adopting a protective tone to monitor the political issues that their own platforms contributed to.

Meta's emphasis on the importance of public debate in turn echoes what Boltanski and Thévenot (2006) analyze as the *civic* order of worth, as well as earlier theories of the public sphere (Habermas, 1989), where private people come together to discuss public concerns. In Habermas' view, the public sphere that emerged in eighteenth-century Europe was characterized by the existence of “rational-critical” discourse supposedly decoupled from private interests (see Fraser, 1990, for a feminist critique). Here, Meta executives praise the role of social media—admittedly owned and operated on an aggressive for-profit basis—as a continuation of the coffee shops that Habermas romanticized.

User Wellbeing

Instead of economic performance or democracy writ large, the third priority turns to the individual as it focuses on fostering and supporting user “wellbeing.” If Meta's public material on democratic debate adopted a paternalistic style, their statements on wellbeing take on a decidedly more caring and maternal tone. In this vein, Meta executives regularly claim to “protect” their “community” and decrease the

“toxic” aspects of social media to ensure the “health” and “happiness” of the platform’s users—all terms laden with multiple and often fuzzy meanings (Gibson et al., 2023).

In his 2012 letter to the SEC, Zuckerberg emphasized the “social mission” of Facebook: “We hope to strengthen how people relate to each other. Even if our mission sounds big, it starts small—with the relationship between two people [. . .]. Relationships are how we discover new ideas, understand our world and ultimately derive long-term happiness” (Ebersman, 2012, p. 67). In this statement, Zuckerberg equates digital connections with strong relationships, happiness, discovery, and openness—all made possible through the platform. This vision determined the labeling of contacts as “friends” on Facebook—a label that has no technical basis, as Bucher (2021, p. 102) reminds us.

It also shaped pronouncements by Facebook executives over the years. For instance, in 2018, Mosseri, then Head of News Feed, described Facebook as a tool to “bring people close together and build relationships.” He was echoing a public post in which Zuckerberg stated that Facebook had been built to “put friends and family at the core of the experience” to “improve our wellbeing and happiness” (Mosseri, 2018). As of 2024, the focus on community wellbeing is still central to the company, as exemplified in the opening sentence of Meta’s Community Standard: “Every day, people use Facebook to share their experiences, connect with friends and family, and build communities” (Meta, n.d.).

Over time, however, the equation of digital connections, community, and happiness became harder to take for granted (Nowland et al., 2018; Sheldon et al., 2011). Users voiced their dissatisfaction with the superficiality of many online “friendships,” reporting negative emotions when seeing their friends’ posts (Aalbers et al., 2019; Shaw et al., 2015). They criticized the oppressive social and beauty standards promoted on social media (Appel et al., 2016). They reported “wasting time” and being “bored” (Gelles-Watnick, 2022). Many of these reactions came up in surveys about Facebook: As early as 2010, user satisfaction was described as “abysmal” (Reuters, 2010).

Facebook ran multiple studies to understand why people felt bad when using the platform. Researchers analyzed the patterns of “emotional contagion” and “social comparison” on Facebook (Kramer et al., 2014); how users defined “meaningful interactions” (Litt et al., 2020); how negative beliefs about Facebook affected users’ perception of time spent on the platform (Ernala et al., 2022); and why users saw Facebook as “problematic” (Cheng et al., 2019). Zuckerberg referred to this body of research in 2018:

Recently we’ve gotten feedback from our community that public content—posts from businesses, brands and media—is crowding out the personal moments that lead us to connect more with each other. [. . .] We feel a responsibility to make sure our services aren’t just fun to use, but also good for people’s wellbeing. [. . .]

The research shows that when we use social media to connect with people we care about, it can be good for our wellbeing. We can feel more connected and less lonely, and that correlates with long term measures of happiness and health. (Zuckerberg, 2018)

To address these issues, Zuckerberg announced a change to the News Feed algorithm to prioritize posts by “friends, family, and groups” (compared to content posted by businesses and news organizations). The change was meant to improve the “wellbeing” of the Facebook “community.” In addition to the redesign of the News Feed, Meta soon unrolled several high-profile features on Instagram. These included messages stating “You’re All Caught Up” in the Instagram Feed (thus putting an end to infinite scrolling; Instagram, 2018); the option to hide the public count of “likes” on Instagram (Instagram, 2021; Meta, 2021a); and the option to set up reminders to “Take a Break” on Instagram.

Interestingly, not all these interventions had noticeable effects on users in terms of improved psychological or physical outcomes. For instance, “Project Daisy”—the pilot program that tested the effects of hiding the public count of “likes” on Instagram—did not make teenagers’ experiences of the platform more positive. Yet Facebook rolled out the change to all users regardless, in part to publicly demonstrate that they took user wellbeing seriously. In an internal discussion leaked to the media in 2021, Meta executives wrote: “A Daisy launch would be received by press and parents as a strong positive indication that Instagram cares about its users” (Wells et al., 2021).

Meta’s public emphasis on user wellbeing endures. In 2023, Mosseri, head of Instagram and support for Threads (designed to compete with X/Twitter), announced that politics and news would not be algorithmically promoted on the new platform. Again, drawing on a well-established trope, he appealed to the “amazing communities” of users and the psychological toll and “negativity” associated with political content to justify the decision:

Politics and hard news are inevitably going to show up on Threads—they have on Instagram as well to some extent—but we’re not going to do anything to encourage those verticals. [. . .] Politics and hard news are important, I don’t want to imply otherwise. But my take is, from a platform’s perspective, any incremental engagement or revenue they might drive is not at all worth the scrutiny, negativity (let’s be honest), or integrity risks that come along with them. There are more than enough amazing communities—sports, music, fashion, beauty, entertainment, etc.—to make a vibrant platform without needing to get into politics or hard news (Mosseri, 2023).

Throughout these statements, Meta executives adopted what Boltanski and Thévenot call a *domestic* order of worth, comparing social media users to dependents having a range of emotional and physical needs that needed to be addressed and cared for (Boltanski & Thévenot, 2006; Bucher, 2021,

p. 210). Taking a step back, this logic of wellbeing bears some similarities with what feminist theorists have analyzed as a “logic of care” (Mol, 2008; Tronto, 2013), in the sense that Meta articulates a sense of responsibility for the nurturing, sustenance, and repair of the relationships and wellbeing of a community of participants. Of course, these statements appear somewhat incongruous given the psychological toll that women and minority groups experience on Facebook and Instagram (Marwick, 2021; Sharp & Gerrard, 2022; Sobieraj, 2020). And indeed, Meta’s efforts toward community wellbeing repeatedly fall short of their stated goals. As we saw with Project Daisy, the public perception of Facebook as “caring” often seemed more important than the actual efficacy of the intended interventions.

Beyond Meta: Logics Across Platforms

So far, our analysis has focused on the case of Meta: We showed how the logics of engagement, public debate, and user wellbeing inductively emerged from Meta executives’ public material, statements, and interviews. To strengthen and expand our argument, we reviewed the public material and executive statements produced by TikTok, Reddit, and X (formerly known as Twitter). We find evidence suggesting that similar discourses operate at these distinct platforms.

In TikTok’s public material, the focus on engagement primarily manifests itself through repeated mentions of “discovery,” “creativity,” and “dialogue”—with explicit mentions of how “fun” the platform is. For instance, the “TikTok for Business” page invites brands to “supercharge [their] TikTok strategy with an *Always Engaged* approach (TikTok, n.d., emphasis added),” stating that “Engagement on TikTok looks different. Here, it’s beyond just likes and shares—it’s this two-way dialogue between brands and audiences that allows brands to grow on TikTok. TikTok’s unique ecosystem of entertainment [. . .] is built on endless discovery, diverse community, and an inclusive culture” (TikTok, n.d.). TikTok also states that it takes community wellbeing seriously: “We care deeply about the well-being of our community members and want to be a source of happiness, enrichment, and belonging” (TikTok, 2023b), adding that “we do not allow content that may put young people at risk of exploitation, or psychological, physical, or developmental harm” (TikTok, 2023c). The public debate angle is a bit less visible, which might be due to TikTok’s ownership structure (ByteDance is a Chinese internet company that also operates Douyin and Toutiao in China). Yet it comes through when the company states that: “In a global community, it is natural for people to have different opinions, but we seek to operate on a shared set of facts and reality” (TikTok, 2023a). In its statement on election integrity, TikTok further discusses the importance of “the informed exchange of civic ideas in a way that fosters productive dialogue” (TikTok, 2023a), which clearly echoes the civic order of worth and the Habermasian public sphere framework.

On Reddit, we find a similar emphasis on engagement, public debate, and wellbeing, but with interesting nuances. For instance, Reddit’s early public material often blends engagement and public debate discourse. Take this 2012 description of the platform on its Frequently Asked Questions (FAQ) page (cited in Gilbert, 2013):

Reddit is a source for what’s new and popular on the web. Users like you provide all of the content and decide, through voting, what’s good and what’s junk. Links that receive community approval bubble up towards #1, so the front page is constantly in motion and (hopefully) filled with fresh, interesting links.

Reddit’s design of “upvotes” was designed to help users identify and make visible good ideas and discussions, yet it has also been found to boost extreme and far-right content, as well as manosphere subreddits (Gaudette et al., 2021; Krendel, 2023). As of 2023, the focus on engagement is clear on the business pages of the company, which emphasize Reddit’s “performance analytics” and its dashboards to “keep a pulse on overall organic engagement” for brands and businesses (Reddit, 2024). The public debate logic remains inherent in Reddit’s goal to empower its users to have diverse conversations while preserving disagreements. For instance, one of the platform’s stated values is to “Keep Reddit Real”: “We don’t understand or agree with everything on Reddit (we’re a vast and diverse group of people, too), and we don’t try to conform Reddit to what we or other people think it should be. We do, though, try to create a space that is as real, complex, and wonderful as the world itself” (spez, 2022). Reddit’s take on community wellbeing uses the same language as its competitors: The company’s public material states that its “mission is to bring community, belonging, and empowerment to everyone in the world, and we do that by keeping Reddit safe, healthy, and real” (Reddit, 2023). Yet they also focus explicitly on authenticity and vulnerability as key values on the platform. As Steve Huffman, Reddit’s co-founder and CEO, notes: “More than any other place on the internet, Reddit is a home for authentic conversation [. . .]. There’s a lot of stuff on the site that you’d only ever say in therapy, or A.A., or never at all” (Isaac, 2023).

Finally, X (formerly Twitter) also features the engagement, wellbeing, and public debate priorities. Since its acquisition by Elon Musk, X has become a fervent defender of a specific version of the public debate discourse—one that centers on free speech absolutism under the tagline of promoting “the exchange of information” (X, 2024b). For instance, a recent company blog post invited users to “Stand with X to protect free speech,” complaining about recent “attacks from activist groups [. . .] who seek to undermine freedom of expression on our platform” (X, 2024d; X Safety, 2023). Interestingly, X explicitly outlines the tension and trade-offs between their free speech values and their commercial bottom line, stating that “X will protect the public’s right to free expression. We will not allow agenda-driven

activists, or even our own profits, to deter our vision.” At the same time, the company seeks to reassure potential advertisers about the centrality of engagement as a goal and set of metrics guiding X’s strategy (X, 2024c). Perhaps unsurprisingly, discourse on wellbeing is currently less visible on the platform’s public pages—but it still exists. For instance, X’s “Abuse and Harassment” statement explains that: “We recognize that if anyone, regardless of background, experiences harassment on X, it can jeopardize their ability to express themselves and cause harm [. . .]. We prohibit behavior and content that harasses, shames, or degrades others. In addition to posing risks to people’s safety, these types of behavior may also lead to physical and emotional hardship for those affected” (X, 2024a). Here the language of harm and health clearly echoes the ideals of the wellbeing logic—even though the reality of how these policies are implemented might differ significantly from their public presentation.

Together, these elements indicate the presence of the logics and repertoires of engagement, wellbeing, and public debate at TikTok, Reddit, and X. This overview in turn surfaces interesting differences. For instance, X under Musk’s ownership seems to prioritize public debate more than any other platform, while TikTok is more vocal about how much it cares about people’s wellbeing.

Discussion

How can we make sense of the revolving priorities of social media platforms and their muddled implementations? In this article, we adopted a meso-level approach and analyzed the different evaluative logics emerging from social media companies’ public material. Focusing on the case of Meta, we inductively identified three distinct priorities: engagement, public debate, and user wellbeing. We outlined the discourses and policies sustaining these different goals and examined how they surfaced at other social media platforms. In this section, we turn to the struggles and tensions that emerge in the daily operations of social media companies, and what this means for the study of online inequality and extraction.

Internal Fractures

Until now, we have examined the logics of engagement, public debate, and wellbeing as if they were operating in separate worlds. Of course, this does not reflect the day-to-day operations of social media companies, which are characterized by ongoing negotiations, conflicts, and trade-offs between people and teams involved in projects that align with these different goals. Due to the secrecy and non-disclosure agreements put in place by social media companies, we do not know much about these internal struggles. Yet, from publicly available material, it appears that most of the time, in these negotiations, engagement comes out on top.

Consider the disconnect between the three following events. First vignette: Zuckerberg, commenting on the 2018

Facebook News Feed algorithm change to promote content by friends and family instead of public pages, explained:

Now, I want to be clear: by making these changes, I expect the time people spend on Facebook and some measures of engagement will go down. But I also expect the time you do spend on Facebook will be more valuable. And if we do the right thing, I believe that will be good for our community and our business over the long term too (Zuckerberg, 2018).

Zuckerberg clearly acknowledged the tension between what we analyze as the logics of engagement (which he refers to as time spent on Facebook and other “measures of engagement”) and user wellbeing. He outlined the distinct temporal frames within which he made sense of this tension, presenting a current (short term) setback in terms of engagement and profitability as an investment in the long-term success of the platform. If we took this statement at face value, we would expect Meta executives to consistently prioritize user wellbeing over other goals.

Second vignette: fast forward to late 2020. Employees inside Facebook realized that their users viewed many of the most viral posts on the platform as “bad for the world.” In response, the employees built a machine learning classifier to downrank such posts, only to have the effort shelved by executives because it reduced engagement metrics. A post published on Facebook’s internal network, leaked to the *New York Times*, explained: “The results were good except that it led to a decrease in sessions [*the number of times users opened Facebook*], which motivated us to try a different approach” (Roose et al., 2020). The message was clear: prioritize wellbeing and public debate goals only insofar as they do not meaningfully cut into engagement.

Third vignette: the 2021 Facebook documents leaked by Frances Haugen revealed multiple cases where the initiatives and recommendations of the Facebook’s Civic Integrity Team were canceled or outranked by Zuckerberg and other top executives. This was so largely because the proposed interventions led to a decrease in engagement metrics, including the ubiquitous “MSI”—a metric that appears on almost every single page of the Haugen files and that served as a North Star indicator for most teams. As one document states about a proposed intervention to protect public discourse, “Mark [Zuckerberg] doesn’t think we could go broad. [. . .] We wouldn’t launch if there was a material trade-off with MSI.” (Brandom et al., 2021; Dwoskin et al., 2021).

The gap between Zuckerberg’s public declarations, which are almost always about Meta’s moral guidance and ethical responsibilities, and the reality of how internal decisions are made is jarring. Facebook’s Civic Integrity Team was clearly tasked with improving and protecting public debate on the platform, while Instagram’s Wellbeing team saw its mission as protecting vulnerable users. Yet in both cases, these teams’ initiatives were blocked or stymied by top management because they were threatening engagement metrics.

These high-level cases give a small preview of the kinds of internal conflicts and negotiations that take place within Meta and other social media companies. They provide a much-needed counterpoint to the companies' public declarations about their commitment to protect democratic integrity and user wellbeing. Based on these examples, it seems that the goals of public debate and wellbeing sit uneasily on the sides of the surveillance and algorithmic apparatus that social media platforms have put in place over the past decades to boost user engagement and sell advertising placement.

Gendered, Racialized, and Extractive Decoupling

Given this dominance of engagement as a North Star goal, why do social media companies even bother hiring the internal teams and putting together the public pages dedicated to improving public debate and user wellbeing?

Our findings evoke what organizational scholars call “decoupling” or “loose coupling,” a process that describes how organizations develop rituals and ceremonial formal structures to gain legitimacy in response to external pressures from civil society and regulators (DiMaggio & Powell, 1983; Friedland & Alford, 1991; Meyer & Rowan, 1977). Given that much of the external pressure coming from the institutionalized environment is ripe with contradictions and ambiguities, organizations tend to “decouple” their public presentation from their internal processes. Decoupling also allows for-profit companies to protect their economic bottom line while maintaining legitimacy. According to neo-institutionalist scholarship, decoupling is more likely in fields and industries where there is some uncertainty about what counts as organizational success (Meyer & Rowan, 1977, p. 354).

In this view, the public debate and wellbeing discourses can be analyzed as “ceremonial props” or smokescreens through which social media companies maintain their public standing and reputation, while their actual guiding logic—engagement—goes unchecked, fostering misinformation and disinformation, hate speech, harassment, and bullying that target vulnerable users. For instance, the public debate priority gained ground after the 2016 U.S. presidential election, during which Facebook was widely criticized for enabling viral misinformation and foreign interference. The wellbeing discourse emerged more gradually in the second half of the 2010s, yet it also maps onto internal company awareness about the growing dissatisfaction of users—especially teenagers—interacting with Facebook and Instagram. Thus, both the public debate and wellbeing discourses can be analyzed as reactive, in the sense that Meta executives sought to anticipate and respond to backlash about their negative impact on democracy and users' mental health.

One could further argue that the visible presence of the public debate and wellbeing discourse and teams is what enables social media companies to save face, receive funding, and continue doing “business as usual,” even when their products are showed to harm and delegitimate users,

reproducing and amplifying racial and gender discrimination (Benjamin, 2019; Noble, 2018; Sobieraj, 2020). This in turn is exactly what Ray (2019, p. 42) analyzes as “racialized decoupling,” in which organizations “decouple formal commitments to equity, access, and inclusion from policies and practices that reinforce, or at least do not challenge, existing racial hierarchies.”

It might also, in Couldry and Mejias' (2019b, p. 341) words, serve “the purpose of ‘social’ media platforms [. . .] to encourage ever more of our activities and inner thoughts to occur on platforms.” The social interactions related to user wellbeing and public debate that occur on platforms constitute a new type of raw material. In this view, wellbeing and public debate content could be part of the “the audacious yet largely disguised corporate attempt to incorporate *all of life* [. . .] into an expanded process for the generation of surplus value” (p. 343, original emphasis).

At the same time, the mere presence of these alternative goals of public debate and community wellbeing may, perhaps, alter internal dynamics within social media companies. As we argued, technology companies are not monoliths: They comprise multiple layers, teams, and categories of actors with different—and often competing—intentions and values (Ali et al., 2023; Bucher, 2021). The existence of goals and discourses legitimating the protection of public debate and user wellbeing within social media companies itself creates frictions and openings. Managers tasked with leading the public debate and wellbeing initiatives often hired employees with distinct qualifications, sociodemographic backgrounds, and political convictions. While this is not in itself a guarantee of social change, it may, under the right conditions, transform how social media companies address the harms they create.

Conclusion

In recent years, social media platforms have voiced paradoxical priorities and executed half measures in response to issues of polarization, misinformation, abuse, harassment, and mental health. To make sense of these inconsistencies, we argue that social media workers and departments are torn between competing evaluative logics and fractured discourses about the company's mission. Focusing on Meta's public material, we identify three key priorities developed by the company to describe and justify its responsibilities: engagement, public debate, and user wellbeing. We find that these distinct evaluative logics are present at other social media platforms—TikTok, Reddit, and X—and that they are often in conflict. Given the centrality of commercial incentives for social media platforms, we note that the public debate and wellbeing logics remain largely subordinate to engagement goals in the companies' day-to-day functioning. This hierarchy contradicts Meta's public declarations but is understood by employees, observers, and regulators, as evidenced by frequent backlashes against the company's inconsistent, decoupled, and contradictory policies. We discuss

how this decoupling between the official discourses of social media companies and what we know of their internal decision-making process allows them to continue doing “business as usual” in the face of massive public critique.

Our analysis also raises additional questions. First, we still know little about the day-to-day functioning of social media companies, largely because of their intense corporate secrecy. We need more in-depth studies of how they make mundane decisions about their products, policies, and personnel. Concretely, this means that scholars need to be given access to internal company processes—not just to deidentified social media data. Second, we need to pay closer attention to how questions of racial and gender justice intersect with social media companies’ stated priorities of improving user wellbeing and the quality of public debate, both of which are theoretically supposed to improve equity and care across social media platforms but may not end up accomplishing these goals. Third, we might consider what social media platforms would look like if they were designed at the outset not to optimize for engagement, but instead to meet the ideals of public debate and user wellbeing. In the meantime, our analysis of social media companies’ public discourses surfaces the complex set of revolving priorities that they claim to uphold, leaving us with the messy work of understanding how these values compete and conflict.

Acknowledgements

This article would not have been possible without the discussions and feedback of the Social Media AI group, including Tatsunori Hashimoto, Michelle Lam, Nathaniel Persily, Tiziano Piccardi, Martin Saveski, Johan Ugander, and Samidh Chakrabarti. We also thank David Stark for his feedback, as well as Rachel Bergmann and Tyler Leeds for fantastic research assistance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: a Hoffman-Yee Research Grant from the Stanford Institute for Human-Centered Artificial Intelligence (HAI).

ORCID iDs

Angèle Christin  <https://orcid.org/0000-0002-7437-2013>

Marijn N. Mado  <https://orcid.org/0000-0002-4491-4117>

Note

1. This comes with broader societal implications. As Van Dijck et al. (2018) argue in *The Platform Society*, social media platforms’ values (described as libertarian, neoliberal, and materialistic) clash with the democratic values safeguarded by most governments of western European countries, including (but not limited to) the right to independent news coverage, worker’s rights, privacy rights, and equal access to education.

References

- Aalbers, G., McNally, R. J., Heeren, A., De Wit, S., & Fried, E. I. (2019). Social media and depression symptoms: A network perspective. *Journal of Experimental Psychology: General*, 148(8), 1454–1462.
- Ali, S. J., Christin, A., Smart, A., & Katila, R. (2023). Walking the walk of AI ethics: Organizational challenges and the individualization of risk among ethics entrepreneurs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 217–226). Association for Computing Machinery. <https://doi.org/10.1145/3593013.3593990>
- Amnesty International. (2022). *Myanmar: Facebook’s systems promoted violence against Rohingya; Meta owes reparations*. <https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebook-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/>
- Ananny, M. (2018, April 4). Checking in with the Facebook fact-checking partnership. *Columbia Journalism Review*. https://www.cjr.org/tow_center/facebook-fact-checking-partnerships.php
- Andersson, H. (2018, July 3). Social media apps are “deliberately” addictive to users. *BBC News*. <https://www.bbc.com/news/technology-44640959>
- Appel, H., Gerlach, A. L., & Crusius, J. (2016). The interplay between Facebook use, social comparison, envy, and depression. *Current Opinion in Psychology*, 9, 44–49.
- Beckert, J., & Aspers, P. (2011). *The worth of goods: Valuation and pricing in the economy*. Oxford University Press.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code* (1st ed.). Polity Press.
- Berry, G., & Taylor, S. J. (2017). Discussion quality diffuses in the digital public square. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1371–1380). <https://doi.org/10.1145/3038912.3052666>
- Boltanski, L., & Thévenot, L. (2006). *On justification: Economies of worth*. Princeton University Press.
- boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brandom, R., Heath, A., & Robertson, A. (2021, October 25). Eight things we learned from the Facebook papers. *The Verge*. <https://www.theverge.com/22740969/facebook-files-papers-frances-haugen-whistleblower-civic-integrity>
- Brock, A. (2020). *Distributed blackness: African American cybercultures*. New York University Press. <https://doi.org/10.18574/nyu/9781479820375.001.00011>
- Bucher, T. (2021). *Facebook* (1st ed.). Polity Press.
- Chander, A., & Krishnamurthy, V. (2018). The myth of platform neutrality. *2 Georgetown Law Technology Review*, 400–416.
- Cheng, J., Burke, M., & Davis, E. G. (2019). Understanding perceptions of problematic Facebook use: When people experience negative life impact and a lack of control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). <https://doi.org/10.1145/3290605.3300429>
- Christin, A. (2020). *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton University Press.
- Clegg, N. (2023, February). Protecting public debate during the war in Ukraine and protests in Iran. *Meta*. <https://about.fb.com/news/2023/02/protecting-public-debate-in-ukraine-and-iran/>

- Cohen, D. (2021, June 30). He created your phone's most addictive feature. Now he wants to build a Rosetta stone for animal language. *GQ*. <https://www.gq.com/story/aza-raskin-interview>
- Couldry, N., & Mejias, U. A. (2019a). *The cost of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- Couldry, N., & Mejias, U. A. (2019b). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, 20(4), 336–349. <https://doi.org/10.1177/1527476418796632>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://doi.org/10.2307/2095101>
- Dixon, S. (2024). *Facebook MAU worldwide 2024*. Statista. <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Dwoskin, E., Newmyer, T., & Mahtani, S. (2021, October 25). The case against Mark Zuckerberg. *The Washington Post*. <https://www.washingtonpost.com/technology/2021/10/25/mark-zuckerberg-facebook-whistleblower/>
- Ebersman, D. (2012, February 1). *Registration statement on form S-1*. [https://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954_10%20\(p.%2077\)](https://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954_10%20(p.%2077))
- Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2022, April 29–May 5). Mindsets matter: How beliefs about Facebook moderate the association between time spent and well-being. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–13). <https://doi.org/10.1145/3491102.3517569>
- Facebook. (2017). Discussion quality diffuses in the digital public square. *Meta Research*. <https://research.facebook.com/publications/discussion-quality-diffuses-in-the-digital-public-square/>
- Facebook. (2023). *Misinformation*. Transparency Center. <https://transparency.fb.com/policies/community-standards/misinformation/>
- Fidelman, M. (2013, October 15). Meet the growth hacking wizard behind Facebook, Twitter and Quora's astonishing success. *Forbes*. <https://www.forbes.com/sites/markfidelman/2013/10/15/meet-the-growth-hacking-wizard-behind-facebook-twitter-and-quoras-astonishing-success/>
- Fligstein, N., & McAdam, D. (2012). *A theory of fields*. Oxford University Press.
- Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, 25/26, 56–80. <https://doi.org/10.2307/466240>
- Frenkel, S., & Kang, C. (2021). *An ugly truth: Inside Facebook's battle for domination*. Harper.
- Friedland, R., & Alford, R. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. In W. W. Powell & P. J. DiMaggio (Eds.) *The new institutionalism in organizational analysis* (pp. 232–263). University of Chicago Press.
- Frier, S. (2020). *No filter: The inside story of Instagram*. Simon & Schuster.
- Gaudette, T., Scrivens, R., Davies, G., & Frank, R. (2021). Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society*, 23(12), 3491–3508. <https://doi.org/10.1177/1461444820958123>
- Gelles-Watnick, R. (2022). *Parents of US teens worry about explicit content, time-wasting on social media*. Pew Research Center. <https://www.pewresearch.org/short-reads/2022/12/15/explicit-content-time-wasting-are-key-social-media-worries-for-parents-of-u-s-teens/>
- Gibson, A. G., Docherty, N. T., & Gillespie. (2023). Health and toxicity in content moderation: The discursive work of justification. *Information, Communication & Society*, 27(7), 1441–1457.
- Gilbert, E. (2013, February). Widespread underprovision on Reddit. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 803–808). ACM Digital Library.
- Gillespie, T. (2010). The politics of "platforms." *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Gleit, N., VP Product, & Social Impact. (2020, August 13). Launching our US 2020 voting information center. *Meta*. <https://about.fb.com/news/2020/08/launching-voting-information-center/>
- Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into the category of bourgeois society*. MIT Press.
- Hao, K. (2021, March). How Facebook got addicted to spreading misinformation. *MIT Technology Review*. <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/>
- Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575–599. <https://doi.org/10.2307/3178066>
- Helft, M. (2008, August 10). Is Google a media company? *The New York Times*. <http://www.nytimes.com/2008/08/11/technology/11google.html>
- Hill Collins, P. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). Routledge.
- Hoffmann, A. L., Proferes, N., & Zimmer, M. (2018). "Making the world more open and connected": Mark Zuckerberg and the discursive construction of Facebook and its users. *New Media & Society*, 20(1), 199–218. <https://doi.org/10.1177/1461444816660784>
- Instagram. (2018, July 2). *Introducing "you're all caught up" in feed*. *Instagram Blog*. <https://about.instagram.com/blog/announcements/introducing-youre-all-caught-up-in-feed>
- Instagram. (2021, May 26). Giving people more control on Instagram and Facebook. *Meta Instagram*. <https://about.instagram.com/blog/announcements/giving-people-more-control>
- Isaac, M. (2023, April 18). Reddit wants to get paid for helping to teach big A.I. systems. *The New York Times*. <https://www.nytimes.com/2023/04/18/technology/reddit-ai-openai-google.html>
- Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1599–1670.
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790. <https://doi.org/10.1073/pnas.1320040111>
- Krendel, A. (2023). The Reddit manosphere as a text and place. In E. Esposito & M. KhosraviNik (Eds.), *Discourse in the digital*

- age: *Social media, power, and society* (pp. 69–88). Taylor & Francis.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38, 201–221.
- Leaver, T., Highfield, T., & Abidin, C. (2020). *Instagram: Visual social media cultures* (1st ed.). Polity Press.
- Lewis, R. (2018). *Alternative influence: Broadcasting the reactionary right on YouTube*. Data & Society.
- Litt, E., Zhao, S., Kraut, R., & Burke, M. (2020). What are meaningful social interactions in today's media landscape? A cross-cultural survey. *Social Media + Society*, 6(3), 1–17. <https://doi.org/10.1177/2056305120942888>
- Marwick, A. E. (2021). Morally motivated networked harassment as normative reinforcement. *Social Media + Society*, 7(2), 1–13.
- Massanari, A. (2017). #Gamergate and the Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3), 329–346.
- McPherson, C. M., & Sauder, M. (2013). Logics in action: Managing institutional complexity in a drug court. *Administrative Science Quarterly*, 58(2), 165–196.
- Meta. (n.d.). *Facebook community standards*. Transparency Centre. <https://transparency.fb.com/en-gb/policies/community-standards/>
- Meta. (2019, February 6). No, your news feed is not limited to posts from 26 friends. <https://about.fb.com/news/2019/02/inside-feed-facebook-26-friends-algorithm-myth/>
- Meta. (2021a, May 26). Giving people more control on Instagram and Facebook. <https://about.fb.com/news/2021/05/giving-people-more-control/>
- Meta. (2021b, May 4). Making emotional health a priority this mental health awareness month. <https://about.fb.com/news/2021/05/mental-health-awareness-month/>
- Meta. (2022a, October 4). How fact-checking works. Transparency Center. <https://transparency.fb.com/features/how-fact-checking-works>
- Meta. (2023). *Transparency center*. <https://transparency.fb.com/features/explaining-ranking/ig-notifications/>
- Meta. (2024). Meta ads. <https://www.facebook.com/business/ads>
- Meta AI. (2019). *Powered by AI: Instagram's explore recommender system*. <https://ai.meta.com/blog/powered-by-ai-instagram-explorer-recommender-system/>
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.
- Mol, A. (2008). *The logic of care: Health and the problem of patient choice*. Routledge.
- Mosseri, A. (2018, January 12). Bringing people closer together. *Meta*. <https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/>
- Mosseri, A. (2023, July 7). *Adam Mosseri (@mosseri) on Threads*. *Threads*. <https://www.threads.net/@mosseri/post/CuZ6opKtHva>
- Napoli, P., & Caplan, R. (2017). Why media companies insist they're not media companies, why they're wrong, and why it matters. *First Monday*, 22(5). <https://doi.org/10.5210/fm.v22i5.7051>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press. <https://doi.org/10.2307/j.ctt1pw9w5>
- Nowland, R., Necka, E. A., & Cacioppo, J. T. (2018). Loneliness and social internet use: Pathways to reconnection in a digital world? *Perspectives on Psychological Science*, 13(1), 70–87.
- Potts, N. (2023, March 29). How meta is advancing democracy. *Meta*. <https://about.fb.com/news/2023/03/how-meta-is-advancing-democracy/>
- Ray, V. (2019). A theory of racialized organizations. *American Sociological Review*, 84(1), 26–53. <https://doi.org/10.1177/0003122418822335>
- Reddit. (2023, June). Transparency report: January to June 2023. <https://www.redditinc.com/policies/2023-h1-transparency-report>
- Reddit. (2024, March 8). Introducing the new toolkit for business growth: Reddit pro is here-upvoted. <https://www.redditinc.com/blog/introducing-the-new-toolkit-for-business-growth-reddit-pro-is-here>
- Reuters. (2010, July 20). Facebook user satisfaction “abysmal.” <https://www.reuters.com/article/urnidgns852573c4006938800025776600370927-idUS218484156520100720>
- Rhodes, L. (Producer), & Orlowski, J. (Director). (2020). *The social dilemma—A Netflix original documentary*. The Social Dilemma. <https://www.thesocialdilemma.com/>
- Roberts, S. T. (2021). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.
- Roose, K., Isaac, M., & Frenkel, S. (2020, November 24). Facebook struggles to balance civility and growth. *The New York Times*. <https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html>
- Rosen, G. (2019, October 21). Helping to protect the 2020 US elections. *Meta*. <https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/>
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1–12. <https://doi.org/10.1177/2053951717738104>
- Sharp, G., & Gerrard, Y. (2022). The body image “problem” on social media: Novel directions for the field. *Body Image*, 41, 267–271.
- Shaw, A. M., Timpano, K. R., Tran, T. B., & Joormann, J. (2015). Correlates of Facebook usage patterns: The relationship between passive Facebook use, social anxiety symptoms, and brooding. *Computers in Human Behavior*, 48, 575–580.
- Sheldon, K. M., Abad, N., & Hinsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: Disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology*, 100(4), 766–775.
- Silverman, C. (2016, November 16). This analysis shows how viral fake election news stories outperformed real news on Facebook. *BuzzFeed News*. <https://www.buzzfeednews.com/article/craig-silverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- Sobieraj, S. (2020). *Credible threat: Attacks against women online and the future of democracy*. Oxford University Press.
- spez. (2022, February 17). *Reddit community values [Reddit post]*. *R/Reddit*. https://www.reddit.com/r/reddit/comments/suvhyq/reddit_community_values/
- Stark, D. (2009). *Heterarchy: The organization of dissonance* (Vol. 245). Princeton University Press.
- Thornton, P. H., & Ocasio, W. (1999). Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958–1990. *American Journal of Sociology*, 105(3), 801–843.
- Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure and process*. Oxford University Press.

- TikTok. (n.d.). *Supercharge your TikTok strategy with an always engaged approach—TikTok for business blog*. TikTok for Business. <https://www.tiktok.com/business/en-US/blog/supercharge-your-tiktok-strategy-with-an-always-engaged-approach>
- TikTok. (2023a, March 8). *Integrity and authenticity*. <https://www.tiktok.com/community-guidelines/en/integrity-authenticity/>
- TikTok. (2023b, March 6). *Mental and behavioral health*. <https://www.tiktok.com/community-guidelines/en/mental-behavioral-health/>
- TikTok. (2023c, March 8). *Youth safety and well-being*. <https://www.tiktok.com/community-guidelines/en/youth-safety/>
- Tronto, J. C. (2013). *Caring democracy: Markets, equality, and justice*. New York University Press.
- Turow, J. (2011). *Media today: An introduction to mass communication*. Taylor & Francis.
- Vaidhyanathan, S. (2018). *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press.
- van Dijck, J., & Poell, T. (2013). *Understanding social media logic* (SSRN Scholarly Paper 2309065). <https://papers.ssrn.com/abstract=2309065>
- van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society*. Oxford University Press. <https://doi.org/10.1093/oso/9780190889760.001.0001>
- Wells, G., Horwitz, J., & Seetharaman, D. (2021, September 14). Facebook knows Instagram is toxic for teen girls, company documents show. *Wall Street Journal*. <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>
- X. (2024a). *Abuse and harassment*. <https://help.twitter.com/en/rules-and-policies/abusive-behavior#:~:text=We%20prohibit%20the%20malicious%2C%20unreciprocated,content%2C%20to%20target%20an%20individual>
- X. (2024b). *Twitter transparency center*. <https://transparency.x.com/en.html>
- X. (2024c). *What are promoted ads?* <https://business.x.com/en/help/overview/what-are-promoted-ads.html>
- X. (2024d). *X's free speech and rights of people—X help*. <https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice>
- X Safety. (2023). *Stand with X to protect free speech*. https://blog.x.com/en_us/topics/company/2023/stand-with-x-to-protect-free-speech
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.
- Zuckerberg, M. (2018, January 12). *Facebook*. https://www.facebook.com/zuck/posts/10104413015393571?ref=embed_post
- Zuckerberg, M. (2020, June 16). Historic Facebook campaign will boost voter registration, turnout and voices. *USA Today*. <https://www.usatoday.com/story/opinion/2020/06/17/facebook-voter-campaign-strengthen-democracy-mark-zuckerberg-column/3191152001/>

Author Biographies

Angèle Christin is Associate Professor of Communication at Stanford University (angelec@stanford.edu)

Michael S. Bernstein is Associate Professor of Computer Science at Stanford University (msb@cs.stanford.edu)

Jeffrey T. Hancock is Professor of Communication at Stanford University (hancockj@stanford.edu)

Chenyang Jia is Assistant Professor in the College of Arts, Media and Design at Northeastern University (c.jia@northeastern.edu)

Marijn N. Mado is a PhD student in communication at Stanford University (mnmado@stanford.edu)

Jeanne L. Tsai is Professor of Psychology at Stanford University (jltsai@stanford.edu)

Chunchen Xu is a postdoctoral scholar in psychology at Stanford University (cxu66@stanford.edu)