Richmond Journal of Law & Technology

Volume 28 | Issue 3 Article 4

4-27-2022

Why Meta Users Need a Public Advocate: A Modest Means to Address the Shortcomings of the Oversight Board

Kevin Frazier
University of California Berkeley School of Law

WHY META USERS NEED A PUBLIC ADVOCATE: A MODEST MEANS TO ADDRESS THE SHORTCOMINGS OF THE OVERSIGHT BOARD

Kevin Frazier [*]

Cite as: Kevin Frazier, *Why Meta Users Need A Public Advocate: A Modest Means To Address The Shortcomings Of The Oversight Board*, 28 RICH. J.L. & TECH. 596 (2021).

[*] Kevin Frazier is a third-year student at the UC Berkeley School of Law. The author is deeply grateful to Professors Kathryn Abrams, Chris Hoofnagle, and Paul Schwartz for fostering his interest in the nexus of technology and the law, to Professor David Eaves and the entire Berkman Klein Center community, and to Evelyn Douek and Quinta Jurecic for their deep dives into content moderation on their Lawfare podcast: "Arbiters of Truth."

ABSTRACT

Meta took an unprecedented step in content moderation when it created an independent board, the Oversight Board, to adjudicate the company's decisions on contested posts. Whether that step constitutes a step forward in the ongoing struggle to moderate online communities depends on whether Meta iterates on its innovation. The current structure of the Oversight Board renders it unable to institute broad and necessary changes to Facebook and Instagram, Meta's main platforms. The creation of an Office of the Public Advocate, charged with representing the interests of Meta's users before the Board, would drastically improve the ability of the Board to positively change Meta's Community Standards. Such an office would serve as a model for other companies to emulate as users seek more influence over their online communities.

I. INTRODUCTION

[1] Meta users have insufficient means to change the Community Standards ("Standards") that govern content moderation on Facebook and Instagram.[1] Meta's existing operations rely on an independent board of experts collectively named the Oversight Board (the "Board"), which is tasked with assisting with the process of enforcing community standards. As it stands, the Board only offers users the means to appeal content decisions made around a user's individual posts, as opposed to content standards for all users. This article proposes that Meta could address this deficiency in its governance structure by creating an Office of the Public Advocate ("OPA"). Through the OPA, users would have a chance to make use of and address their concerns through the Board's processes and procedures.[2]

[2] In its current form, the Board has limited control over its docket and lacks the ability to change Meta's Standards.[3] The second part of this article outlines the current content moderation process for Facebook and Instagram and identifies some of its flaws and shortcomings. With the creation of the OPA, the experts sitting on the Board should have the authority to assess the merits of OPA arguments made on behalf of users, then require Meta to make the Community Standards comply with the Board's assessments. This process would improve the information ecosystems of Meta's current and

[1] *How we update the Facebook Community Standards*, META (Jan. 20, 2022), https://transparency.fb.com/policies/improving/deciding-to-change-standards/ [https://perma.cc/UYJ2-4P9S].

[2] *How to appeal to the Oversight Board*, META (Jan. 19, 2022), https://transparency.fb.com/oversight/appealing-to-oversight-board/ [https://perma.cc/PW6F-2LUS].

[3] *Oversight Board recommendations*, META (Jan. 14, 2022), https://transparency.fb.com/oversight/oversight-board-recommendations/ [https://perma.cc/FG9F-KUGB]; *Appealing Content Decisions on Facebook or Instagram*, OVERSIGHT BD., https://oversightboard.com/appeals-process/ [https://perma.cc/5J25-CBG6].

future platforms.[4] The third part of this article examines other contexts in which a public advocate has advanced the interests of disaggregated communities like Meta and how a public advocate could benefit both Meta and its users. Indeed, users would benefit from the ability to express concerns that reflect regional and context-specific moderation issues. Meta would benefit by implementing an independent content moderation process that takes Meta's business priorities into account while also placing important and time-sensitive policy decisions into appropriately-trained hands.

[3] Though this article focuses on an improvement to Meta's content moderation structure, many other social media platforms would benefit from the utilization of public advocates to represent their users. As of 2021, fewer than thirty percent of all adults have "a lot or some trust" in the information that comes from social media.[5] This figure is down from thirty-four percent in 2016.[6] Dwindling trust may explain why, in recent years, people seem more willing to leave social media platforms. After years of bad press, Meta's Facebook platform experienced a drop-off in users,

[4] Georgia Wells, *How People Can Make Smarter—and Healthier—Social-Media Choices*, WALL ST. J. (Dec. 13, 2021, 10:00 AM), https://www.wsj.com/articles/smarter-healthier-social-media-choices-11639177212 [https://perma.cc/LK6D-NVWQ] ("Facebook's own research found about one in eight of the app's users reported engaging in compulsive use of the company's app that affected their sleep, work, parenting or relationships.").

[5] Jeffrey Gottfried & Jacob Liedke, *Partisan divides in media trust widen, driven by a decline among Republicans*, PEW RSCH. CTR. (Aug. 30, 2021), https://www.pewresearch.org/fact-tank/2021/08/30/partisan-divides-in-media-trust-widen-driven-by-a-decline-among-republicans/ [https://perma.cc/R9QR-9ZJ7] ("About a quarter of Americans (27%) say they have at least some trust in the information that comes from social networking sites, with just 4% expressing that they have a lot of trust in it.").

[6] *Id.*

especially among young Americans.[7] Based on an internal Meta memo about their flagship platform, The Verge reported that, as of 2021, "[t]eenage users of the Facebook app in the US had declined by 13 percent since 2019 and were projected to drop 45 percent over the next two years."[8] More generally, researchers like Dr. Brian Primack, Professor of Public Health and Medicine and Dean of the College of Education and Health Professions at the University of Arkansas, have identified a growing discomfort across society with the "double-edged sword" that is social media.[9] In an article exploring how people can use social media in a more healthy manner, Dr. Primack noted broad recognition of the fact that social media "can breed feelings of depression and anxiety, and even isolation and hatred."[10] As more individuals question and leave platforms after detecting the negative aspects of social media use, platforms have incentive to reform their content moderation processes and policies. The fourth part of this article summarizes the importance of internal modifications to Meta's content moderation strategy because of insufficient external remedies and, finally, addresses counterarguments to the use of a public advocate.

[7] Megan Leonhardt, *Teens have been losing interest in Facebook for years, internal and external data shows*, FORTUNE (Oct. 25, 2021, 3:19 PM), https://fortune.com/2021/10/25/facebook-teens-usage-harm-studies/ [https://perma.cc/X8XJ-837Z].

[8] Alex Heath, *Facebook's Lost Generation*, THE VERGE (Oct. 25, 2021, 7:00 AM), https://www.theverge.com/22743744/facebook-teen-usage-decline-frances-haugen-leaks [https://perma.cc/GB3K-NZ7Y].

[9] Wells, *supra* note 4; *see also* BRIAN A. PRIMACK, YOU ARE WHAT YOU CLICK: HOW BEING SELECTIVE, POSITIVE, AND CREATIVE CAN TRANSFORM YOUR SOCIAL MEDIA EXPERIENCE 138–141 (2021).

[10] Wells, *supra* note 4.

II. THE CURRENT PROCESS

[5] Social media platforms have a range of processes meant to moderate content and reduce the spread of hate speech, defamatory content, other illegal content, and content that violates the platform's community standards.[11] There is no universal way to moderate content because each platform has its own objectives.[12] The determination of what constitutes acceptable activity is based on the laws of the jurisdictions where the platforms operate and the preferences of their respective users.[13]

[6] Meta's existing process struggles to consider the diversity inherent to billions of unique users.[14] Indeed, Meta's Standards guide content moderation at Meta.[15] These Standards apply to all Facebook users and determine what content is allowed on the platform.[16]

---

[11] *See* Merlene Leano, *Social Media Moderation Guidelines*, NEW MEDIA SERVS. (Jan. 15, 2021), https://newmediaservices.com.au/social-media-moderation-guide/ [https://perma.cc/6MF5-UZUW].

[12] *See id.* (comparing different platforms' common problems and describing different types of social media moderation).

[13] *Id.*; *see* Roman Leal, *Can the Facebook Supreme Court "say what the law is?": The Limits of Oversight Board Sovereignty*, YALE CYBER LEADERSHIP FORUM (July 20, 2021), https://cyber.forum.yale.edu/blog/2021/7/20/can-the-facebook-supreme-court-say-what-the-law-is-the-limits-of-oversight-board-sovereignty [https://perma.cc/LHG6-TGQ7] (describing how local laws impact Oversight Board decisions).

[14] META, *supra* note 1.

[15] *Id.*

[16] *Id.* (outlining the set of standards for Facebook—Instagram applies a different but similar set of standards).

[7] Generally, as is true at Meta, platform employees, rather than users, exercise the most control over specific community content standards. Meta claims its Standards reflect "feedback from people and the advice of experts in fields like technology, public safety and human rights."[17] However, Meta's Standards, shielded from alteration by the Oversight Board, are meant to decide which content remains on the platform to maintain a balance of "free speech and safety."[18]

[8] Governments occasionally force changes to social media platform standards by passing new laws or exerting informal influence in the direction of a certain change.[19] Twitter, for example, ceded to pressure from governments in the E.U. when it shifted its policy to remove accounts promoting terrorism.[20] Users may have been given a chance to effectuate change if Twitter implemented public advocates to represent user's interests.

[9] Whereas most companies rely solely on employees to enforce their community standards, Meta created an independent board of experts to

[17] Facebook Community Standards, META, https://transparency.fb.com/policies/community-standards [https://perma.cc/LZP2-5Z4N].

[18] *See Just the Facts on the Oversight Board*, META, https://about.facebook.com/actions/oversight-board-facts [https://perma.cc/LNV3-4MAG].

[19] *See generally, e.g.*, Megan McKnelly, *Untangling SESTA/FOSTA: How the Internet's "Knowledge" Threatens Anti–Sex Trafficking Law*, 34 BERKELEY TECH. L.J. 1239 (2019) (discussing how a recently passed law required increased moderation by social media platforms).

[20] *See generally* Natasha Lomas, *Tech giants' slowing progress on hate speech removals underscores need for law, says EC*, TECH CRUNCH (Oct. 7, 2021, 11:19 AM), https://techcrunch.com/2021/10/07/tech-giants-slowing-progress-on-hate-speech-removals-underscores-need-for-law-says-ec/ [https://perma.cc/9ZF2-4RF3].

assist with the process.[21] Named the Oversight Board, the Board reviews decisions made by the Meta content moderation process to determine whether content should stay on the platform.[22] Meta acknowledged the need for the Board because it does not think it should make content moderation decisions in its sole discretion.[23] However, the employees at Meta charged with specifying the company's Standards retain ultimate authority over its content because those employees can unilaterally modify the Standards that the Board is obliged to apply when reviewing a decision to take down content.[24]

[10] Companies are aware that they need to have more objective reviews of their standards to ensure that profits are not the primary motivating factor in determining their policies.[25] For example, consider that the causal effect of social media inciting violence and problematic behavior has grown stronger in recent years.[26] Left unchecked, these posts can have deleterious effects on political stability, human rights, and, more crudely, public perceptions of the platform. The use of Facebook to "foment division and incite offline violence" in Myanmar, as acknowledged by Facebook,

[21] META, *supra* note 18.

[22] *Id.*

[23] *Id.*

[24] *Id.*

[25] *See* Nina Brown, *Regulatory Goldilocks: Finding the Just and Right Fit for Content Moderation on Social Platforms*, 8 TEX. A&M L. REV. 451, 454 (2021).

[26] *Id.* at 454–55 (detailing how Facebook, Twitter, and YouTube have "lenient" stances on issues such as hate speech that have allowed groups to organize online and bring about physical harm).

demonstrates the physical effects of virtual vitriol.[27] Independent analysts determined that the failure of Facebook employees to catch these posts "helped to fuel modern ethnic cleansing" in the Southeast Asian country, as summarized by Alexandra Stevenson of the *New York Times*.[28] Social media also continues to play a more central role in politics and has been regarded as a sort of public square.[29] However, social media platforms have largely been left to self-regulate. This tendency for self-regulation stems from a general understanding that "overregulation at the request of—or inducement by—the government is inherently more problematic than a platform's own decision to over-remove content."[30]

[11] The processes and procedures designed to facilitate platform self-regulation lack meaningful ways for the users to participate in the formation of the community standards that govern user content. Changes to community standards represent a promising way to alter the information ecosystems created by each platform.[31] Algorithms serve as the first means of enforcing community standards. Indeed, algorithms are designed to

[27] Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. TIMES (Nov. 6, 2018), https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html [https://perma.cc/R5B2-JDVC] (quoting a Facebook executive).

[28] *Id.*

[29] Preventing Online Censorship, 85 Fed. Reg. 34,079, 34,081 (June 2, 2020); See also U.S. DEP'T.OF JUST., SECTION 230 – NURTURING INNOVATION OR FOSTERING UNACCOUNTABILITY? 21 (2020), https://www.justice.gov/file/1286331/download [https://perma.cc/L5BN-8LDV] ("Unconstrained discretion is particularly concerning in the hands of the biggest platforms, which today effectively own and operate digital public squares.").

[30] Brown, *supra* note 25, at 476.

[31] *Id.* at 480 (describing the current method of self-regulation and the failure of platforms to be transparent about the community standards used to govern moderation).

automatically scan posts for content that violates community standards.[32] However, algorithms cannot enforce these standards perfectly. Instead, "algorithms consider only what is being said, paying little regard to the post's purpose or what it *actually* communicates to the platform's audience."[33] Although humans are flawed, they have the critically important role in moderation of catching the nuances that algorithms overlook.[34] But even with robust training, human moderators make errors.[35] Specifically, human moderators "struggle with difficult decisions and apply community standards inconsistently—a product of vague guidelines, broad discretion, and their own subjective biases."[36]

[12] The vast majority of content decisions are made by the algorithm trained to apply the community standards.[37] The sheer volume of posts per day on sites like Facebook and Instagram necessitate reliance on such automated mechanisms.[38] However, it is possible that algorithms assessing social media platform community standards become less effective as the quantity of content grows; social media algorithms assess the use of language, and, "language is incredibly complicated, personal, and context dependent, which limits the algorithm's abilities to differentiate between

[32] *Id.* at 477–79.

[33] *Id.*

[34] *Id.*

[35] Brown, *supra* note 25, at 479.

[36] *Id.*

[37] *Cf. Id*. at 477 (explaining how major social networks such as Facebook, Twitter, and YouTube all employ algorithms as the primary filter in their algorithm and human "mixed approach" systems to content moderation).

[38] *Id.* at 456.

permissible and problematic posts."[39] Both algorithms and human moderators have their downfalls. Platforms with global reach must find better ways to adjudicate individual and multi-user claims related to the enforcement and content of community standards.

[13] In Meta's case, the Board, created to respond to user challenges to content moderation decisions, only evaluates whether the algorithms and humans appropriately applied the pre-existing Standards.[40] Cases before the Board follow a simple process. In the even that Meta, via an automated review of a post, a human review, or a mix of both,[41] decides to take down or keep up a post, the author of the post may ask Meta to review their decision.[42] If Meta affirms their original decision, the user can then appeal to the Board. Next, the Board decides whether to take the user's case. The Board makes these decisions based on three factors: the number of users affected by the issue presented in the case, whether the case addresses issues related to public discourse, and whether the case raises meaningful questions about Meta's policies.[43] Finally, the Board issues a binding decision with respect to that post.[44] In theory, how the Board decides one

[39] *See id.*

[40] Brown, *supra* note 25, at 491–92.

[41] *See, e.g.,* Issie Lapowsky, *Meta will let some users know when their posts are removed by AI*, PROTOCOL (Mar. 1, 2022), https://www.protocol.com/bulletins/meta-automated-content-moderation-alert [https://perma.cc/N2RG-6DA2] (detailing how Meta uses automated processes as well as analog processes to review posts).

[42] META, *supra* note 18.

[43] Elin Hofverberg, *Facebook's New "Supreme Court" – The Oversight Board and International Human Rights Law*, LIBRARY CONG. (Mar. 16, 2021), https://blogs.loc.gov/law/2021/03/facebooks-new-supreme-court-the-oversight-board-and-international-human-rights-law/ [https://perma.cc/Z2XP-T86M].

[44] META, *supra* note 18.

case should inform future content moderation decisions in similar instances, but there's no requirement for Meta to follow the Board's recommendations.[45]

[14] Meta's ability to ignore the Board's recommendations may lead to outcomes that are preferable to the platform, rather than the outcome that users prefer. The business models of these social media platforms critically rely on a "combination of [Section 230-based] immunity and lack of regulatory oversight."[46] For these business models to work, users "freely creat[e] and upload[] content . . . with little risk of liability for the publisher of that content—the platform."[47] Section 230 shields platforms from liability for sharing certain content under certain circumstances.[48] This means that Facebook and other platforms have avoided liability despite hosting socially-disfavored content such as revenge porn, acts of egregious violence, and race-based attacks.[49] As long as profit objectives are prioritized in the formation of community standards,[50] users will likely experience an information ecosystem that promotes content which is most likely to go viral, draw advertisers, retain user attention, and ultimately make money for the platform.

---

[45] *Id.* (distinguishing between Board *decisions* on particular cases, which are binding, and Board *recommendations* on broader Meta policy, which are not).

[46] *See* Brown, *supra* note 25, at 454.

[47] *Id.*

[48] *Id.*

[49] Alina Selyukh, *Section 230: A Key Legal Shield For Facebook, Google Is About To Change*, NPR (Mar. 21, 2018, 5:11 AM), https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change [https://perma.cc/HD77-JW3C].

[50] *See* Brown, *supra* note 25, at 454.

[15] Additionally, the Board's consideration of single posts, rather than classes of posts, also diminishes the value of any recommendations for changes to the Standards. The Board cannot *sua sponte* identify a group of posts or even a single post that it believes merits review. Instead, the Board only has the option of reviewing posts that (1) have been removed or kept up by Meta, then (2) had that decision challenged by the authoring user, and, finally, (3) had that decision affirmed by Meta.[51] The fact-specific nature of Meta's decision on a single post constrains the extent to which a decision by the Board can have broader implications. Consider a case concerning hate speech posted to Facebook, decided by the Board in 2021, that illustrates this point.[52] In this case, the Board reviewed whether to affirm Meta's decision to remove a post made by a Facebook user in a public group discussing "'multi-racialism' in South Africa,"[53] where the user "argued that poverty, homelessness, and landlessness have increased for black people in the country since 1994."[54] In the user's post, the user referred to people who disagreed with the user's stance by using terms identified by Meta as "prohibited slurs for the Sub-Saharan market."[55] The Board affirmed Meta's decision to remove the post by relying on Facebook's Hate Speech Community Standard, which "prohibit[s] the use of slurs targeted at people based on their race, ethnicity and/or national origin."[56] In being even more specific, the Board articulated several reasons

[51] *See* META, *supra* note 18.

[52] *Oversight Board upholds Facebook Decision: Case 2021-011-FB-UA*, OVERSIGHT BD. (Sept. 2021), https://oversightboard.com/news/404712621226343-oversight-board-upholds-facebook-decision-case-2021-011-fb-ua/ [https://perma.cc/469Q-39JA].

[53] *Id.*

[54] *Id.*

[55] *Id.*

[56] *Id.*

for affirming the post's removal; the user chose to use "the most severe terminology possible in the country," the severity of which is magnified by the "legacy of apartheid" in South Africa.[57] Additionally, the Board agreed with Facebook in that, although Facebook permits, in some cases, content that "condemn[s] or raise[s] awareness of the use" of hate speech, this user decidedly did not qualify for that exception.[58] The Board then recommended "greater transparency around Facebook's slur list" in order to improve "procedural fairness" in the enforcement of Meta's Hate Speech policy.[59]

[16] It is not clear how Meta content moderators should apply this decision to similar cases. For the moderator to follow this ruling, will they have to consult experts on what qualifies as "the most severe terminology" in that region or see evidence of an ongoing domestic struggle over racial issues? The recommendations fail to offer guidance on the scope of the decision because they merely offer broad areas for further investigation by Meta.

[17] Just as enforcing the speed limit can only do so much to prevent traffic fatalities, ensuring the proper application of community standards can only do so much to improve the information ecosystem of a social media platform. Eventually, changing the speed limit may be the best option. To successfully motivate a speed limit change, perhaps community members contact their local representatives personally or voice their concerns in a more public forum. Hopefully, the jurisdiction would hear those complaints and, if determined to be valid, would establish a new, safer speed limit. Social media users do not have such an option. Even if users leave comments, write op-eds, and rally broad support, the Board has no authority

[57] *Oversight Board*, *supra* note 52.

[58] *Id.*

[59] *Id.*

to adjudicate the strength of their argument nor to enforce any policy recommendation.[60] An OPA could provide users with someone to advocate for desired change. An OPA could also provide the Board with a representative to advance the public's arguments in a hearing alongside the platform's own arguments. This additional step in the content moderation scheme would have a far greater chance of improving the information ecosystem of platforms because it could better achieve balance between free speech and safety.

III. WHAT IS THE ROLE OF A PUBLIC ADVOCATE?

[18] Public advocates exist in several adjudicatory and regulatory settings. Private entities generally lack someone tasked with representing the interests of the public in company processes. [61] However, ombudspersons within such entities play a role akin to that of public advocates though they only raise the concerns of employees.[62] The authority and mission of public advocates varies based on the priorities of the public and of the regulatory bodies in front of which the advocates advance those priorities. A brief examination of various missions and policies that public advocate offices pursue reveals a variety of options for the creation of a

[60] *Oversight Board: Further asked questions*, META (Jan. 19, 2022), https://transparency.fb.com/oversight/further-asked-questions/ [https://perma.cc/C5ET-9QLZ] (" [Question:] Are the Oversight Board's recommendations binding? [Answer:] Unlike the board's content decisions on individual cases, recommendations are not binding. Meta is committed to both considering these recommendations as important input on our internal policy processes and publicly responding to each recommendation within 30 days.").

[61] *See The new private-sector ombudsmen*, Policy Options (Nov. 1, 2003), https://policyoptions.irpp.org/magazines/corporate-governance/the-new-private-sector-ombudsmen/ [https://perma.cc/DYG2-75RT] (describing the increase in private entities creating ombudspersons tasked with handling complaints from employees).

[62] *Id.*

public advocate tasked with assisting Meta's users. These options are then applied to Meta's content moderation structure.

A. Examples of Public Advocate Offices and the Benefits They Bring

[19] Some public advocates have a meaningful role in proactively resolving conflicts identified by the public. For example, the Citizen Public Advocate in Snohomish County, Washington, defines their mission as "help[ing] [to] find transparent solutions for the people of Snohomish County."[63] Accomplishing this mission requires lending "a guiding hand in providing a path for ethical accountability within county government."[64] In practice, the Citizen Advocate is an "independent entity [that] serves as an impartial intermediary between the citizen and general Snohomish County government."[65] In this role, the Citizen Advocate can launch investigations into complaints.[66] Though the Citizen Advocate does not resolve those complaints, it can outline suggested resolutions for all involved parties.[67]

[20] Other public advocates have narrower missions. The California Public Utilities Commission (CPUC) includes the Public Advocates Office

[63] *Citizen Public Advocate*, SNOHOMISH CNTY. WASH., https://snohomishcountywa.gov/2352/Citizen-Public-Advocate [https://perma.cc/6BDA-E6ZF] (quoting Jill McKinnie).

[64] *Id.*

[65] *Id.*

[66] *Id.*

[67] Public Advocate Annual Report, SNOHOMISH CNTY. (Dec. 31, 2016), https://snohomishcountywa.gov/3574/Public-Advocate-Report [https://perma.cc/2633-M3EN].

("PAO").[68] This office's mission is to "advocate for the lowest possible monthly bills" for consumers of the state's regulated utilities.[69] The PAO's responsibility is to represent and advocate for customer interests in hearings before the CPUC.[70] To achieve this goal, the PAO gathers information on the regulated utilities for public consumption, petitions for rule changes, and issues fact sheets and other reports to keep the public up to date.[71] The PAO claims that this work resulted in millions of dollars in savings in 2021 for Californians.[72]

[21] Finally, some public advocates focus on promoting specific communities. For example, prior to Governor Chris Christie eliminating the office in 2010, New Jersey's executive branch included a Department of the Public Advocate ("Department").[73] The Department focused primarily on the needs of the elderly, people with mental illness or developmental disabilities, consumers, and children.[74] By refocusing on relationship building, the Department "change[d] the legal landscape on an important

[68] *About Public Advocates Office*, CAL. PUB. UTILS. COMM'N, https://www.publicadvocates.cpuc.ca.gov/ [https://perma.cc/JD2Q-2AGE].

[69] *Id.*

[70] *Id.*

[71] *Id.*

[72] *Id.* ("The Public Advocates Office saved consumers $3.7 billion dollars in lower utility revenues and avoided rate increases.").

[73] N.J. Rev. Stat. § 52:27EE-86 (2013); *See also* Tom Johnson, *Public Advocate Office Quietly Headed for Elimination*, NJ SPOTLIGHT NEWS (May 10, 2010), https://www.njspotlightnews.org/2010/05/10-0507-1647/ [https://perma.cc/NNG6-HC4H].

[74] *Id.*

issue facing New Jersey."[75] In some cases, the Department went so far as to sue other government agencies to get positive results for the public.[76] The Department's work resulted in meaningful benefits that improved the lives of every New Jersey resident.[77] In fact, the Department may have been uniquely positioned to accomplish goals such as reducing rates of lead poisoning in children, an issue that is difficult for any other official, agency, or stakeholder to identify.[78] Each of these examples of public advocate offices show the benefits of having an office that exclusively advocates for the public.

B. How a Public Advocate Could Work in Meta's Content Moderation Structure

[22] Platforms have wide discretion over how to moderate content.[79] The designs of their respective systems often reflect their own profit motives by striving to create a safe online space and by taking content moderation seriously.[80] However, it is also true that companies have consciously

[75] *See id.* at 51; Johnson, *supra* note 73 (quoting former public advocate Ronald Chen).

[76] DEP'T OF THE PUB. ADVOCATE, 2008 ANNUAL REPORT (2009), https://dspace.njstatelib.org/xmlui/bitstream/handle/10929/19056/2008.pdf?sequence=1&isAllowed=y [https://perma.cc/3WJ6-YT9J].

[77] Johnson, *supra* note 73 ("If it had not been for the public advocate, Potter said, the principle of establishing a right of access to New Jersey beaches might not exist, Public Service Enterprise Group might have built floating nuclear power plants off the coast, and the Supreme Court might not have ordered every town to provide affordable housing."

[78] *See id.*

[79] Brown, *supra* note 25, at 480.

[80] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1626–27 (2018).

designed their platforms to foster a less healthy information ecosystem based on their profit motives, rather than designing the platform according to the demands of their users.[81] In the deluge of documents unleashed by Frances Haugen, a former Facebook employee turned whistleblower, investigators uncovered evidence that the algorithm used to determine what content appeared on users' Facebook newsfeeds amplified hate speech.[82] A change to Facebook's algorithm resulted in reshared posts, which disproportionately contain "[m]isinformation, toxicity, and violent content," having greater reach—ending up on the News Feeds of more users by virtue of the algorithm more heavily weighting content that had been reshared.[83] Haugen's disclosures also included information about Instagram. According to Haugen and others, Instagram's algorithm has even tragically resulted in teenagers' "desire to kill themselves" due to continued exposure to algorithmically-amplified content related to self-harm.[84] The Board currently has no means to specify and enforce changes to the algorithm related to content selection and amplification.

[23] Meta's response in the wake of the Haugen revelations about the algorithms directing content on Facebook and Instagram show the need for a different approach to content moderation. Instagram, responding to increased attention, created new tools for users, especially for younger

[81] *Id.* at 1627.

[82] Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead.*, WALL ST. J. (Sept. 15, 2021, 9:26 AM), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [https://perma.cc/2MYE-9JAV].

[83] *Id*.

[84] Georgia Wells et al., *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, WALL ST. J. (Sept. 14, 2021, 7:59 AM), https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739 [https://perma.cc/8ATK-9VFP].

users, to reduce their exposure to potentially harmful content and to spend less time on the app.[85] Experts do not expect these changes to result in substantial changes among users, though Instagram cites internal studies that indicate their efficacy.[86] Though these changes do not represent the entirety of internal responses to Haugen's disclosures, stakeholders have demanded more action beyond what has been taken.[87] In response to this heightened interest in content moderation, Congress has hosted hearings, proposed bills, and made statements centered around regulating social media.[88] However, none of the bills have gained traction.[89]

[24] If an OPA existed, a portion of users (the threshold could be a fixed number or percentage of users taking some affirmative action to support the case) on Facebook, Instagram, or any future Meta platform within the Board's bailiwick could have fought for an explicit change to Meta's practices immediately after the disclosure. Those users could complete a form or otherwise indicate their desire for the OPA to pursue the case. Alternatively, the OPA could pursue the case on its own volition unless a fixed number or percentage of users directed the OPA not to do so. In this

[85] Shara Tibken, *Instagram Unveils New Teen Safety Tools Ahead of Senate Hearing*, WALL ST. J. (DEC. 7, 2021, 10:28 AM), https://www.wsj.com/articles/instagram-unveils-tools-to-keep-teens-safe-including-parental-controls-11638864001 [https://perma.cc/6T8Y-6MHY].

[86] *See id*. (noting that experts doubt that opt-in tools will result in much change but that Instagram has reported that users embrace such tools).

[87] *See id.*

[88] *Id*.; *see, e.g.*, Diane Bartz, *U.S. senators announce bipartisan social media data transparency bill*, REUTERS (Dec. 9, 2021, 5:05 AM), https://www.reuters.com/world/us/us-senators-announce-bipartisan-social-media-data-transparency-bill-2021-12-09/ [https://perma.cc/CDH5-RF72].

[89] *See* Bartz, *supra* note 88 (pointing out that the Senate bill lacks a companion bill in the House of Representatives).

latter case, the OPA could prompt the Board to consider policy changes to address an increase in misinformation related to vaccines, for example, unless a certain percentage of Facebook or Instagram users objected to the OPA bringing such a case. Whatever the mechanism, the OPA would ideally articulate some suggestion for the Board to improve the platform. Indeed, Facebook made several changes, including a "major overhaul . . . to its News Feed algorithm" to reduce misinformation, toxicity, and violent content, and "boost 'meaningful social interactions'" on the platform.[90] However, those changes had an undesirable effect from Meta's perspective: they resulted in less user engagement with Facebook.[91] The OPA in that matter could have asked the Board to host a hearing on the best way to implement the aforementioned changes to make public interest a priority.

[25] The OPA would argue for the algorithm that incorporated those changes and could consult third-party stakeholders to present the Board with a broader range of information from which to make their decisions, akin to the submission of an amicus brief. In this case, the OPA could ask researchers like Noah Giansiracusa, a mathematics professor at Bentley University, to summarize and share their findings on the Facebook algorithm. Giansiracusa's brief, for instance, might suggest that, "limiting 'deep reshares' (where content is reshared not only by the original poster's network of friends or followers, but also their friends' friends, and their friends' friends' friends, and so on)," could effectively curtail the spread of disinformation.[92] Facebook would then have a chance to present its own argument in the hearing. In the OPA system envisioned below, the Board

[90] Hagey & Horwitz, *supra* note 82.

[91] *Id.*

[92] Kaleigh Rogers, *Facebook's Algorithm is Broken. We Collected Some Suggestions on How to Fix It*, FIVETHIRTYEIGHT (Nov. 16, 2021, 6:00 AM) (quoting Noah Giansiracusa) https://fivethirtyeight.com/features/facebooks-algorithm-is-broken-we-collected-some-spicy-suggestions-on-how-to-fix-it/ [https://perma.cc/6TZZ-MRJU].

could then publish a decision with binding effect as to which algorithm Facebook would have to use moving forward.[93]

[26] As this hypothetical hearing makes clear, the OPA system would require several changes to the current content moderation structure. The new system would necessitate an expansion of the Board to manage the increased number of cases it would need to hear. Additionally, the expanded Board would likely need to create intermediate boards that operate at the regional level. The current content moderation structure pays shockingly little attention to the 90% of Facebook users that reside outside of the U.S. and Canada and the billions of users that speak languages other than English.[94] By creating regional public advocate offices and boards to hear cases by those branches of the OPA outside of the U.S. and Canada, Meta could begin to address its failure to reduce violations of its Standards in the majority of countries it serves. The implementation of regional offices would also empower users to engage with a process that is more responsive to their unique cultural and linguistic needs.

[27] The OPA itself would need substantial resources to realize its mission. Regional public advocate offices would include lawyers, researchers, and user liaisons. Meta created an independent trust to support

[93] *But see* Grimes v. Donald, 673 A.2d 1207, 1214 (Del. 1996) (holding that "[d]irectors may not delegate duties which lie 'at the heart of the management of the corporation,'" which may limit the ability for Facebook to delegate its power to an OPA. Such corporate law issues are beyond the scope of this paper) (internal citations omitted); Del. Code. Ann. tit. 8, § 141(a) ("the business and affairs of every corporation organized under this chapter shall be managed by or under the direction of a board of directors . . .").

[94] *See*, FACEBOOK, FB EARNINGS PRESENTATION Q3 2021 2, https://s21.q4cdn.com/399680738/files/doc_financials/2021/q3/FB-Earnings-Presentation-Q3-2021.pdf [https://perma.cc/TD2W-BDUC].

the Board.[95] A similar approach could provide sufficient and reliable funding for the OPA.

[28] The mission of the OPA would also need to be specified. As outlined previously, the specificity of missions varies among public advocates.[96] One potential mission for the Meta OPA would be the reduction of hate speech, mis- and dis-information, and toxic speech.[97]

[29] The mechanics for how a Board decision in favor of the OPA would result in changes to Meta's Standards also needs to be explored. One mechanism that would preserve Board discretion involves the OPA presenting the Board with a menu of changes, including Standards changes. If the Board sides with the OPA, Meta would have to institute change. The Board could retain even more discretion by having the authority to set timelines for changes. With the help of the OPA, the Board could monitor whether the changes result in the outcomes desired by the users and the OPA. If a change to the Standards did not result in the desired outcome, then the Board, again with assistance of the OPA, could test another of the OPA's suggested policies.

[30] Most importantly, the Board would need the authority to mandate changes to Meta's Standards and procedures. This change would divest significant authority from Meta employees. This change is necessary in

[95] *See* Elizabeth Culliford, *Facebook Pledges $130 Million to Content Oversight Board, Delays Naming Members*, REUTERS (Dec. 12, 2019, 11:09 AM), https://www.reuters.com/article/us-facebook-oversight/facebook-pledges-130-million-to-content-oversight-board-delays-naming-members-idUSKBN1YG1ZG [https://perma.cc/HQV3-WPNW].

[96] *See supra* Part III.

[97] Lynne Tirrell, *Toxic Speech: Toward an Epidemiology of Discursive Harm*, 45 UNIV. OF ARK. PRESS 139, 140 (2017).

light of Meta employees' conscious decisions to uphold policy that is likely to produce unhealthy information ecosystems.[98]

[31] However, some negative effects are possible with this proposed change. By forcing the Board (and possibly intermediate level boards) to take cases related to the concerns of many users, there is a chance that resulting community standards will be more restrictive of speech. The current process can only make changes on a post-by-post basis because it only reviews Meta's decision with respect to a single post. This means that even the most aggressive decision by the Board would affect future posts that mirror a specific fact pattern. Consequently, Meta's platforms will continue to host a wide range of speech. If, however, the Board could consider classes of posts, such as posts about vaccine misinformation, then the Board could forbid more kinds of speech. Such power could hinder the caliber and diversity of discourse on Meta platforms. Indeed, a change needs to be made. Even though permitting the OPA to bring broad cases before the Board may carry risks, the status quo has resulted in the problematic amplification of socially-deleterious content.

IV. WHY ALTERNATIVE MEANS TO REGULATE FACEBOOK HAVE FAILED

[32] Officials on both sides of the political aisle have advocated for reforming Section 230 as a way to address issues such as disinformation.[99] However, the discussed reforms are likely to make matters worse. First, any universal content moderation rules will likely fail because of the distinct goals of each platform and the unique characteristics of their respective

[98] *See* Hagey & Horwitz, *supra* note 82.

[99] *See* Cristiano Lima, *Can Congress unite on Section 230 reform? This top Democrat has hope.*, WASH. POST (Dec. 1, 2021, 9:14 AM), https://www.washingtonpost.com/politics/2021/12/01/can-congress-unite-section-230-reform-this-top-democrat-has-hope/ [https://perma.cc/2VRG-68CB].

users.[100] Slight changes to Section 230 would lack the nuance required to address the idiosyncratic nature of each platform. Increased government control over content moderation decisions could result in the most proactively self-regulated platforms scaling back their processes and procedures to comply with the government's approach.[101]

[33] The judicial branch will likely also fail to produce policy that addresses content moderation issues. Decades of precedent have tied the hands of jurists.[102] Additionally, case law precedent has provided broad immunity to internet companies, evoking a sense of worry amongst some rule makers that companies like Meta are disincentivized to block objectionable content.[103] It is also unclear whether jurists have the training and familiarity required to formulate rules that would improve the current information ecosystem.

---

[100] *See* Emma Llansó, *Platforms Want Centralized Censorship. That Should Scare You*, WIRED (Apr. 18, 2019, 9:00 AM), https://www.wired.com/story/platforms-centralized-censorship/ [https://perma.cc/97RQ-ZZSA].

[101] *See* Brown, *supra* note 25, at 483.

[102] Daniel Holznagel, *Enforcing the Rule of Law in Online Content Moderation: How European High Court decisions might invite reinterpretation of CDA § 230*, AMERICAN BAR ASSOCIATION (Dec. 9, 2021), https://www.americanbar.org/groups/business_law/publications/blt/2021/12/online-content-moderation/ [https://perma.cc/Q7XY-2BX6] (Summarizing how American courts have broadly interpreted section 230 and other relevant laws to grant social media platforms "wide discretion to self-regulate via Community Standards.")

[103] *See, e.g.*, Daisuke Wakabayashi, *Legal Shield for Social Media is Targeted by Lawmakers*, N.Y. TIMES (Dec. 15, 2020), https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html [https://perma.cc/E938-7VPN] (identifying that many cases brought against content moderators are quickly dismissed due to Section 230's broad legal protections).

[34] Users are not without fault in the creation of unhealthy information ecosystems. Dr. Primack points out that humans have natural vulnerabilities that attract them to certain negative content and lure them toward certain behaviors.[104] Of course, Facebook's algorithm would have no hate speech to spread if users did not generate that content. Even so, platforms cannot shift the blame for unhealthy information ecosystems fully onto their users. Hate speech will likely always exist, but no single human has the ability to quickly and easily make it spread around the world without the aid of a social media platform. If Facebook and other platforms do serve as public forums, they should have an obligation to do more than merely accept and amplify the worst of human behavior.

V. CONCLUSION

[35] The OPA system outlined in Part Three, or something that closely resembles that structure, offers a chance to make Standards more responsive to the needs and realities of global users. The current pursuit of a balance between free speech and safety has drifted off course. Neither Meta nor Congress has proven capable of getting the platform back on track. The Office of the Public Advocate could be the structural addition that brings about the necessary balance to internet content moderation.

[36] Whether or not Meta creates a public advocate, the company should continue to test iterations of its oversight structure. The company's exploration of the metaverse, "an embodied internet where you're in the experience, not just looking at it,"[105] will result in more content to moderate. Under the current structure, the Board has a limited ability to assist with

[104] *See* Wells, *supra* note 4 (interviewing Dr. Primack).

[105] Shirin Ghaffary & Sara Morrison, *Can Facebook monopolize the metaverse?,* VOX (Feb. 16, 2022, 6:30 AM), https://www.vox.com/recode/22933851/meta-facebook-metaverse-antitrust-regulation [https://perma.cc/W6XM-BEYT].

Meta's goal to create spaces in which users "socialize, learn, collaborate and play in ways that go beyond what we can imagine."[106]

[37] The world has never known an entity with as many stakeholders as Meta. What has worked to moderate conversations in town halls and foster community among hundreds, thousands, and millions of people may not scale to a community of billions. Meta has the resources, the brain power, and the obligation to think creatively about how its governance structures can foster more than profit.

[106] *Welcome to Meta*, META, https://about.facebook.com/meta/ [https://perma.cc/JCA2-4PYB].