

**THE META OVERSIGHT BOARD AND THE EMPTY PROMISE
OF LEGITIMACY**

*Evelyn Douek**

ABSTRACT

The Meta Oversight Board is an audacious experiment in self-regulation by one of the world's most powerful corporations, set up to oversee one of the largest systems of speech regulation in history. In the few years since its establishment, the Board has in some ways defied its many skeptics, by becoming a consistent and accepted feature of academic and public discourse about content moderation. It has also achieved meaningful independence from Meta, shed light on the otherwise completely opaque processes within the corporation, instantiated meaningful reforms to Meta's content moderation systems, and provided an avenue for greater stakeholder engagement in content moderation decision-making. But the Board has also failed to live up to core aspects of its role, in ways that have gone underappreciated. The Board has consistently shied away from answering the hardest and most controversial questions that come before it — that is, the very questions it was set up to tackle — and has not provided meaningful yardsticks for quantifying its impact. Understanding why the Board eschews these questions, and why it has nevertheless managed to acquire a significant amount of institutional legitimacy, suggests important lessons about institutional incentives. Ultimately, this Article argues, the current political environment incentivizes a kind of oversight that is formalistic and unmoored from substantive goals. This is a problem that plagues regulatory reform far beyond the Board itself, and shows that generalized calls for “more legitimate” content moderation governance are underspecified and may encourage poor outcomes.

* Assistant Professor, Stanford Law School. Many thanks to Anupam Chander, Ros Dixon, Thomas Kadri, Genevieve Lakier, Clare Mathias, Martha Minow, and Rory Van Loo. All oversights my own.

TABLE OF CONTENTS

ABSTRACT	373
TABLE OF CONTENTS.....	374
I. INTRODUCTION.....	374
II. THE BOARD’S BEGINNING AND ENDS	381
<i>A. The Design</i>	382
<i>B. The Objectives</i>	384
III. REVIEWING THE BOARD’S PERFORMANCE	390
<i>A. The Board’s Success Stories</i>	391
1. Fulfilling its Basic Functions	392
2. The Board’s Assertions of Independence	397
3. Prompting Reforms to Meta’s Systems.....	401
4. Making Meta More Transparent and Accountable.....	406
5. Making Content Moderation More Democratic	408
6. Legitimacy Building.....	412
<i>B. The Board’s Formalistic Approach</i>	414
1. Rulings Without Reason.....	414
2. Formalistic Measures of Impact.....	425
3. Persistent Strategic Avoidance.....	430
IV. LEGITIMACY’S EMPTY PROMISE.....	434
<i>A. The Importance of Institutional Incentives</i>	438
<i>B. The Revealed Preference for Performative Governance</i>	441
<i>C. Sociological Legitimacy as a Poor Marker of Success</i>	443
V. CONCLUSION.....	444

I. INTRODUCTION

Over the past four years, a single institution has adjudicated whether the President of the United States should be able to use one of his preferred channels of communication with an audience of over thirty-five million people,¹ how much weight should be given to the United Kingdom Metropolitan Police’s assessments of the dangers of certain music,² whether COVID-19 misinformation should be

1. Former President Trump’s Suspension Case, 2021-001-FB-FBR (Oversight Bd. May 5, 2021) [hereinafter *Trump Suspension Case*], <https://www.oversightboard.com/decision/FB-691QAMHJ/> [<https://perma.cc/PHY7-4MGM>] (upholding Facebook’s decision to “restrict then-President Donald Trump’s access to posting content on his Facebook page and Instagram account”).

2. UK Drill Music Case, 2022-007-IG-MR (Oversight Bd. Nov. 22, 2022) [hereinafter *UK Drill Music Case*], <https://www.oversightboard.com/decision/IG-PT5WRTLW> [<https://perma.cc/C4JL-8B9Y>].

suppressed online,³ and how to deal with inflammatory and threatening statements from the Prime Minister of Cambodia.⁴ The same institution has decided disputes that touch on some of the world's most contentious subjects, from conflict between Israel and Palestine,⁵ the invasion of Ukraine,⁶ and the pervasiveness of gender-based violence.⁷ It has engaged with some of the hardest, most consequential questions about speech regulation in the modern world — questions about which there is profound, intractable societal disagreement. And it has done all of this despite a complete lack of formal legal authority.

The institution is the Meta (née Facebook)⁸ Oversight Board (the “Board”), an institution that sits in kind of a twilight zone: being both a decision-maker exercising immensely consequential power and a made-up body that exists at the whim of its creator. Its “case decisions” look like court rulings, but they technically bind no one.⁹ It purports to interpret and apply state-created international human rights law (“IHL”), but it has no state-given mandate to do so.¹⁰ The Board's rulings have impact in the real world, and yet its power is somewhat a fiction. Created voluntarily by Meta as a form of self-regulation, the Board's decisions have force only insofar and for as long as Meta — the company the Board was set up to hold accountable — agrees they do.

And yet, despite this significant handicap on its authority, the Board has received a remarkable degree of attention and respect in the academic and media discourse around content moderation. Its decisions are regularly reported in leading media outlets like *The New York*

3. Removal of COVID-19 Misinformation Policy Advisory Opinion, PAO-2022-01 (Oversight Bd. Apr. 20, 2022), <https://www.oversightboard.com/decision/PAO-SABU4P2S/> [<https://perma.cc/2NW3-RD9G>].

4. Cambodian Prime Minister Case, 2023-003-FB-MR (Oversight Bd. June 29, 2023), <https://www.oversightboard.com/decision/FB-6OKJPNS3/> [<https://perma.cc/FL3U-8ZAU>].

5. Shared Al Jazeera Case, 2021-009-FB-UA (Oversight Bd. Sept. 14, 2021), <https://www.oversightboard.com/decision/FB-P93JPX02/> [<https://perma.cc/98VM-8ARS>].

6. Russian Poem Case, 2022-008-FB-UA (Oversight Bd. Nov. 16, 2022), <https://www.oversightboard.com/decision/FB-MBGOTVN8/> [<https://perma.cc/FEE4-XBFW>].

7. Violence Against Women Case, 2023-002-IG-UA, 2023-005-IG-UA (Oversight Bd. June 12, 2023), <https://www.oversightboard.com/decision/IG-H3138H6S/> [<https://perma.cc/6DL2-F57H>].

8. For simplicity, this Article will refer throughout to the parent company of Facebook and Instagram as “Meta,” even when referring to documents or events that preceded the company's formal name change from “Facebook.”

9. See *infra* Section II.A.

10. See *infra* Section III.B.1.

Times,¹¹ *The Washington Post*,¹² *Reuters*,¹³ and *The Guardian*.¹⁴ This news coverage of the Board resembles, in tone and content, reporting on judicial decisions or administrative agencies. “Meta Oversight Board calls for Cambodian leader’s accounts to be suspended” reads one headline, for example.¹⁵ “Instagram told to reinstate music video removed at request of Met police” says another.¹⁶ Reporters often treat the Board, in other words, as a meaningful source of authority whose decisions are news that its readers should know about.

It’s not only the media that consistently and meaningfully engages with the Board in ways that both suggest and bestow a degree of legitimacy. Numerous academics and civil society groups have submitted public comments to the Board with respect to particular cases, including the ACLU,¹⁷ the United Nations Special Rapporteur on Minority Issues,¹⁸ the International Commission of Jurists,¹⁹ the Electronic Frontier Foundation,²⁰ and many other leading institutions and individuals

11. E.g. Maya King, *Instagram and Facebook Should Update Nude Photo Rules, Meta Board Says*, N.Y. TIMES (Jan. 20, 2023), <https://www.nytimes.com/2023/01/20/technology/meta-nudity-standards-overhaul.html> [<https://perma.cc/X82V-KTK3>].

12. E.g. Naomi Nix, *Oversight Board Tells Meta to Restore Post Comparing Russians to Nazis*, WASH. POST (Nov. 16, 2022), <https://www.washingtonpost.com/technology/2022/11/16/facebook-oversight-board-ukraine-war/> [<https://perma.cc/TEH4-74B4>].

13. E.g. Katie Paul, *Meta’s Oversight Board Tells Company to Allow “Death to Khameinei” Posts*, REUTERS (Jan. 9, 2023), <https://www.reuters.com/technology/met-oversight-board-tells-company-allow-death-khameinei-posts-2023-01-09/> [<https://perma.cc/3WA4-4TTY>].

14. E.g. Alex Hern, *Instagram Told to Reinstate Music Video Removed at Request of Met Police*, GUARDIAN (Nov. 22, 2022), <https://www.theguardian.com/technology/2022/nov/22/instagram-told-to-reinstate-music-video-removed-at-request-of-met-police> [<https://perma.cc/UEJ5-W4GM>].

15. Regine Cabato & Rebecca Tan, *Meta Oversight Board Calls for Cambodian Leader’s Accounts to Be Suspended*, WASH. POST (June 29, 2023), <http://www.washingtonpost.com/world/2023/06/29/meta-hun-sen-cambodia-suspended-accounts/> [<https://perma.cc/2LCZ-U3XP>].

16. Hern, *supra* note 14.

17. Emma Llansó & Aliya Bhatia, *CDT & ACLU Urge Meta Oversight Board to Protect Speech in Abortion-Related Cases*, CTR. FOR DEMOCRACY & TECH. (June 29, 2023), <https://cdt.org/insights/cdt-aclu-urge-meta-oversight-board-to-protect-speech-in-abortion-related-cases/> [<https://perma.cc/L4RL-6D52>]; Daphne Keller, *Daphne Keller and ACLU File Comment to Meta Oversight Board in “UK Drill Music” Case*, STAN. FREEMAN SPOGLI INST. (Aug. 23, 2022), <https://fsi.stanford.edu/news/daphne-keller-and-aclu-file-comment-uk-drill-music-case> [<https://perma.cc/SU53-U39Q>].

18. U.N. Special Rapporteur on Minority Issues, Public Comment by U.N. Special Rapporteur on Minority Issues Relating to Cases on Hate Speech and Minorities (Dec. 8, 2020), <https://www.ohchr.org/en/statements/2020/12/public-comment-un-special-rapporteur-minority-issues-relating-cases-hate-speech> [<https://perma.cc/JVH2-T9T3>].

19. *Cambodia: ICJ Submits Public Comment to Oversight Board on the Case of Prime Minister Hun Sen’s Violent Speech on Facebook*, INT’L COMM’N OF JURISTS (Mar. 29, 2023), <https://www.icj.org/cambodia-icj-submits-public-comment-to-oversight-board-on-the-case-of-prime-minister-hun-sen-violent-speech-on-facebook/> [<https://perma.cc/KZQ6-UTBM>].

20. Jillian C. York, *EFF and ECNL’s Comment to the Meta Oversight Board on the Term “Shaheed”*, ELEC. FRONTIER FOUND. (Apr. 24, 2023), <https://www.eff.org/deeplinks/2023/04/eff-and-ecnls-comment-meta-oversight-board-term-shaheed> [<https://perma.cc/M7PJ-3M87>].

in the digital governance space and beyond.²¹ Government actors have done the same: American lawmakers have submitted comments to the Board,²² as have the U.K. Metropolitan police.²³ The Israeli Supreme Court has referenced the Board in one of its judgments as an example of due process that Meta offers its users.²⁴ The United Nations High Commissioner for Human Rights has approvingly noted the Board's work in its annual report.²⁵ The Board has also established academic cachet. It has been written about extensively.²⁶ Its decisions are included alongside judicial opinions and legislative materials in the most widely used legal databases, Westlaw and Lexis+.²⁷ The Board's decision in the *Trump Suspension Case* is included in a leading Internet law casebook,²⁸ and another opinion has even been the subject of a case comment in the *Harvard Law Review*.²⁹

While this degree of attention to, and acceptance of, the Board might one day seem completely unremarkable, it's important to underline for readers from that future that this was far from inevitable when the Board was created. Early reactions to Meta CEO Mark Zuckerberg's decision to create the Board in 2018 were extremely skeptical.³⁰ Many suggested that the Board should not be engaged with at all, because to do so was either fruitless or, worse, would necessarily

21. OVERSIGHT BD., 2022 ANNUAL REPORT 55–57 (2022) [hereinafter 2022 ANNUAL REPORT] (listing various prominent individuals who submitted comments).

22. See, e.g., Ashley Gold, *Republicans Raise Bias Claims to Board Reviewing Trump's Facebook Ban*, AXIOS (Feb. 11, 2021), <https://www.axios.com/2021/02/11/republicans-raise-bias-claims-to-board-reviewing-trumps-facebook-ban> [https://perma.cc/P9UY-6863]; Press Release, Rep. Lori Trahan, Trahan, Schiff Urge Meta to Continue Removing Dangerous Covid-19 Misinformation (Aug. 3, 2022), <https://trahan.house.gov/news/document/single.aspx?DocumentID=2578> [https://perma.cc/A6GE-8BJL]; Cecilia Kang, *Lawmakers Slam Facebook Oversight Board's Decision to Uphold Trump Ban*, N.Y. TIMES (May 5, 2021), <https://www.nytimes.com/2021/05/05/technology/facebook-oversight-board-decision-reaction.html> [https://perma.cc/N7G9-VANA].

23. *UK Drill Music Case*, *supra* note 2.

24. HCJ 7846/19 Adalah Legal Center for Arab Minority Rights in Israel v. State Attorney's Office – Cyber Department ¶ 8 (2021) (Isr.).

25. OFF. HIGH COMM'R FOR HUM. RTS., UNITED NATIONS HUMAN RIGHTS REPORT 2021 at 58 (2021), <https://www2.ohchr.org/english/OHCHRreport2021/> [https://perma.cc/MP9S-8GXJ].

26. See *infra* Section III.A.6 and articles cited throughout.

27. Oversight Board decisions appear under “Administrative Decisions & Guidance” in Westlaw and “Administrative Materials” in Lexis+. Search “FB Oversight Bd” in *Administrative Materials* on Westlaw, or “Facebook Oversight Board” in *Administrative Decisions & Guidance* on Lexis+.

28. JAMES GRIMMELMAN, INTERNET LAW: CASES & PROBLEMS 577–89 (12th ed. 2022).

29. Case Comment, *Oversight Board Finds a Facebook Rule's Application Violates International Human Rights Law: Case Decision 2021-004-FB-UA*, 135 HARV. L. REV. 1971 (2022).

30. FACEBOOK, GLOBAL FEEDBACK & INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS 10 (2019) (“Early feedback ranged from considered skepticism to cautious optimism, and raised fundamental questions about the purpose, scope, operationalization, and impact of the proposed body . . .”).

legitimize a corrupt institution. The Board was described as a “folly,”³¹ “dysfunctional by design,” and “toothless[;]”³² “a McGuffin;”³³ having a “Potemkin quality;”³⁴ and many other similar things in print — and probably some much worse off the record. Such skepticism is not surprising: the Board was the unilateral creation of a corporate, profit-driven entity and was given a narrowly defined remit and no coercive legal power.³⁵ These are not typically the characteristics that inspire unmitigated confidence or widespread public respect for a decision-making authority. As such, the Board started with a very low baseline of sociological legitimacy, and understandably so.

Yet only a few years later, skepticism of the Board has receded. Far from being dismissed as a mere sham, the Board’s existence is largely taken for granted, and its work is taken seriously.³⁶ This Article asks how this change came about, whether the shift in the Board’s public and academic reception is justified by what the Board has achieved so far, and what lessons this holds for social media governance more generally. Understanding how the Board has come to occupy the position it does matters, not only because the Board oversees what is by volume one of the most expansive speech forums in history, but also because the Board experiment sheds light on broader dynamics in content moderation governance. How stakeholders have engaged with the Board reveals what they want and expect from such governance institutions. And as regulators, companies, and the public alike search for ways to bring greater oversight to content moderation systems, the Board holds both salutary lessons and cautionary tales for these institutional designers.

On the one hand, and contrary to many people’s expectations, the Board has not been entirely toothless. It has chafed against the extremely limited remit Meta originally gave it, pushed the limits of its authority, and in doing so sought out broader impact on Meta’s systems than merely deciding individual cases.³⁷ As a result, the Board has brought about meaningful reforms at Meta, some of which civil society

31. Siva Vaidhyanathan, *Facebook and the Folly of Self-Regulation*, WIRED (May 9, 2020, 2:58 PM), <https://www.wired.com/story/facebook-and-the-folly-of-self-regulation/> [<https://perma.cc/YG4R-U7N8>].

32. Jessica J. González & Carmen Scurato, *Everyone on Facebook’s Oversight Board Should Resign*, WIRED (Apr. 17, 2021, 9:00 AM), <https://www.wired.com/story/opinion-everyone-on-facebooks-oversight-board-should-resign/> [<https://perma.cc/CA98-GDVD>].

33. Roger McNamee & Maria Ressa, *Facebook’s “Oversight Board” Is a Sham. The Answer to the Capitol Riot Is Regulating Social Media*, TIME (Jan. 28, 2021, 10:30 AM), <https://time.com/5933989/facebook-oversight-regulating-social-media/> [<https://perma.cc/5RSD-78RC>].

34. Kevin Roose, *Facebook’s ‘Supreme Court’ Tells Zuckerberg He’s the Decider*, N.Y. TIMES (May 6, 2021), <https://www.nytimes.com/2021/05/06/technology/facebook-oversight-board-trump.html> [<https://perma.cc/P9TH-333W>].

35. See *infra* Section II.A.

36. See *infra* Section III.A.6.

37. See *infra* Section III.A.2.

and activists had been seeking for years to no avail.³⁸ And it has shed light on some of the otherwise entirely opaque systems that make millions of speech decisions every day.³⁹ It has been rewarded by becoming in many ways part of the platform governance establishment.

But the Board has also underdelivered in underappreciated ways and failed to fulfill some of its most important tasks. First, despite its design centering the Board's role of public reason-giving, the Board has not explained the normative framework that guides its thinking about online speech governance.⁴⁰ This is a huge missed opportunity — the central normative difficulty of content moderation is how to adapt principles created to constrain *governmental* power over expression to the very different context of *private* content moderation systems. This task is hard but vital in an age where corporate power over everyday speech has never been greater. There is little in modern public discourse that is not affected by the decisions that platforms make — everything is a content moderation problem. Thus, developing a normative foundation for speech rights that goes beyond protection from governmental interference is essential if free speech principles are to be made meaningful in the platform era. The Board is uniquely placed to do this work as one of the most high-profile experiments in content moderation governance, made up of free expression experts who have access to an unending supply of content moderation cases to review and significant resources at their disposal. And yet, the Board has largely eschewed the normative questions that underpin its role.⁴¹ It has adopted IHRL as the basis for its decisions, but its application of IHRL to Meta's platforms has been shallow and provides little insight into what understanding of free speech guides its work.⁴² This undermines not only the predictability of the Board's decisions, but also the Board's role as a source of public reasoning that explains and justifies the rules that govern a major segment of the online public sphere.

Second, if the Board's normative contributions to content moderation debates have been underwhelming, its material impacts on online speech are altogether harder to quantify. The task of quantification has been made more difficult by the Board's emphasis on metrics that are easier to measure but poor proxies for actual impact. The Board tracks Meta's implementation of its decisions and recommendations, but it does not track the more important question of whether and how those changes actually improve people's lived experience and the exercise of their rights.⁴³ In short, the Board has emphasized formalism in the

38. *See infra* Section III.A.3.

39. *See infra* Section III.A.4.

40. *See infra* Section III.B.1.

41. *See infra* Section III.B.1.

42. *See infra* Section III.B.1.

43. *See infra* Section III.B.2.

changes it seeks from Meta and adopted a simplistic approach to measuring the results.

Third, in high-profile or especially hard cases, the Board has too often ducked its responsibility to act as the decider of last resort and avoided giving a definitive answer altogether. This responsibility might have been thought to be the *raison d'être* of the Board. But instead, the Board has found ways to be as uncontroversial as possible, even when deciding the most controversial cases. The result has been to prolong, rather than resolve, some of the most important and high-stakes arguments about online speech rules.

The Board's failure to deliver in these important ways raises questions about why the Board has adopted its current approach, and why it has succeeded in garnering increasing sociological legitimacy regardless. This puzzle is the core question that this Article ultimately seeks to answer. What it shows is that "legitimacy," although often invoked as the lodestar of successful governance, is an ambiguous indicator of what we might want from governance institutions, and its pursuit can create perverse institutional incentives. This insight is critical not only for the Board but for platform governance institutional design in general.

This Article proceeds in three Parts. Part II describes the beginning and ends of the Board. It briefly sets out the history of the Board's creation and the design choices of its creators, and it shows how the pursuit of "legitimacy" was core to the Board's purpose. The Board was fundamentally a response to a growing perception that Meta's content moderation practices were illegitimate, and that the exercise of such incredible power over so much speech should not be so unaccountable. Therefore, while stakeholders articulated many things that they hoped the Board would do — push back against pressure to remove ever more content online, reduce Meta's power (and responsibility), provide an avenue for more democratic input in content moderation governance, bring greater accountability and transparency to Meta's decision-making, provide public reasoning and rationales for Meta's most consequential and controversial decisions — ultimately, all these goals were in service of making Meta's content moderation "more legitimate."

Part III then engages in a performance review. It canvasses the Board's successes and the ways in which the Board has delivered on some of its promise as an institution. Then it turns to the Board's shortfalls and the ways it has engaged in a formalistic kind of governance that fails to provide firm foundations for, or strong evidence of the benefits of, its work. This Part shows that while the Board has achieved much more than many expected, it has also underdelivered in ways that are both significant and underappreciated.

Part IV then suggests an explanation for why the Board shies away from core parts of its role, and why this has not (yet) had reputational

costs. Ultimately, it argues, what the Board's (albeit qualified) success suggests is that the lack of consensus as to the goals for effective content moderation governance creates space for institutions to define success in ways that may suit them but may have little benefit for others. As long as the Board seems better than the available alternatives for decision-making (in this case, completely opaque and unaccountable power ultimately resting in the hands of Mark Zuckerberg) and exhibits some features of "good governance," it appears to be a valuable improvement over the status quo. Indeed, it *is* an improvement. But that is a low bar, and we should demand more from the institutions that oversee some of the most important, and most expansive, speech forums in history. As an exercise in the revealed preferences of stakeholders in social media governance, then, the story of the Board is somewhat discouraging. While the Board has had some successes, it has ultimately so far carried out a form of governance that prioritizes form over substance. The danger is not that the Board's choices are "illegitimate," or seen as such. Instead, the danger is that they are seen as legitimate and disconnected from any other outcome goals we might think content moderation governance should pursue, and that these perceptions influence the design and actions of all content moderation governance, far beyond the Board.

II. THE BOARD'S BEGINNING AND ENDS

The origins of the Board reflect a pervasive anxiety about the "legitimacy" of content moderation governance. The Board was created at a moment when long-standing concerns about the unaccountable power platforms wielded over the digital public sphere had reached a crescendo, and the Board was Meta's attempt to placate these concerns. This Part briefly describes these origins and the design of the Board,⁴⁴ before turning to examine its goals and how it was intended to meet them. What this story shows is that the mechanisms by which the Board would improve content moderation governance were always somewhat vaguely articulated. But the core contribution that Meta hoped the Board would make was always clear: build trust and legitimacy for Meta's decision-making about the difficult, contestable, and

44. Readers interested in the story of the creation of the Board will find comprehensive accounts elsewhere. See, e.g., Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 *YALE L.J.* 2418 (2020); Kate Klonick, *Inside the Making of Facebook's Supreme Court*, *NEW YORKER* (Feb. 12, 2021), <https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court> [<https://perma.cc/7FNH-RBFG>]; Thomas E. Kadri, Response, *Juridical Discourse for Platforms*, 136 *HARV. L. REV. F.* 163, 169–85 (2022); Evelyn Douek, *Facebook's "Oversight Board": Move Fast with Stable Infrastructure and Humility*, 21 *N.C. J.L. & TECH.* 1, 9–46 (2019).

consequential online speech decisions that the company has to make every day.

A. *The Design*

From the beginning, it was obvious that the Board was an attempt by Meta to shape public discourse. Meta’s CEO Mark Zuckerberg first publicly floated the idea of “some sort of structure, almost like a Supreme Court” for his company in a podcast interview in April 2018.⁴⁵ This was a little over a week before Zuckerberg appeared at the first of the many congressional hearings in which lawmakers lambasted him about Meta’s decision-making,⁴⁶ and the proximity of these two events is surely no coincidence. As the timing suggests, the Board is a creature born of the techlash,⁴⁷ and understanding this context sheds light on the particular public pressures to which Meta was responding in creating it. In 2018, Meta was busy trying to contain the damage to its reputation caused by revelations about Russian influence operations on its platforms during the 2016 U.S. presidential elections, the Cambridge Analytica scandal, and a growing general sentiment that tech platforms had become too big, too powerful, and too unaccountable.⁴⁸ As pressure on the company continued to mount, Zuckerberg formally announced in a blog post in November 2018 that Meta would create an independent body that would hear user appeals against the company’s content moderation decisions,⁴⁹ kicking off a process that culminated in the first Board members being announced in May 2020.⁵⁰

The final blueprint of the Board depicted a body intended to comprise forty members,⁵¹ who have a maximum of three three-year

45. Ezra Klein, *Mark Zuckerberg on Facebook’s Hardest Year, and What Comes Next*, VOX (Apr. 2, 2018, 6:00 AM), <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge> [<https://perma.cc/54ZU-4LJ2>].

46. See *Mark Zuckerberg Testimony: Senators Question Facebook’s Commitment to Privacy*, N.Y. TIMES (Apr. 10, 2018), <https://www.nytimes.com/2018/04/10/us/politics/mark-zuckerberg-testimony.html> [<https://perma.cc/TE8F-ZHST>].

47. See generally Eve Smith, *The Techlash Against Amazon, Facebook, and Google — and What They Can Do*, ECONOMIST (Jan. 20, 2018), <https://www.economist.com/briefing/2018/01/20/the-techlash-against-amazon-facebook-and-google-and-what-they-can-do> [<https://perma.cc/NW7T-7DFQ>].

48. See Katy Steinmetz, *Congress Never Wanted to Regulate Facebook. Until Now*, TIME (Apr. 12, 2018, 6:41 AM), <https://time.com/5237432/congress-never-wanted-to-regulate-facebook-until-now/> [<https://perma.cc/PDY3-6UAM>].

49. Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (Nov. 15, 2018), <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/> [<https://perma.cc/UJ57-93A4>].

50. See Nick Clegg, *Welcoming the Oversight Board*, META (May 6, 2020), <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/> [<https://perma.cc/Y5TT-D2FP>].

51. OVERSIGHT BD., OVERSIGHT BOARD CHARTER, art. 1, § 1 (2019) [hereinafter CHARTER], https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf [<https://perma.cc/5VN4-N89X>]; OVERSIGHT BD., OVERSIGHT BOARD BYLAWS, art. 1, § 1.4

terms.⁵² (In practice, the Board has never had anywhere near forty members — it has had a maximum of twenty-three members to date.)⁵³ These members hear appeals from users or take referrals from Meta itself.⁵⁴ The Board’s Charter (the “Charter”) states that the Board should “determine whether [Meta’s content enforcement decisions are] consistent with [Meta’s] content policies and values.”⁵⁵ In doing so, the Board should “pay particular attention to the impact of removing content in light of human rights norms protecting free expression.”⁵⁶

Despite the name “Oversight Board,” which might be thought to imply sweeping oversight of Meta generally, the jurisdiction of the Board — the kinds of decisions and cases it is empowered to review — has always been extremely limited and explicitly does not cover some of the most consequential parts of the company’s operations.⁵⁷ Within the ambit of the Board’s authority, though, the Charter states that the resolution of each case (that is, the decision about the individual piece of content in question) is “binding” (although no enforcement mechanism is specified), and that Meta should implement it promptly, both with respect to the individual piece of content at issue in the case and also with respect to “identical content with parallel context.”⁵⁸ In every decision, the Board may also include a policy advisory statement making recommendations for Meta’s future policy development.⁵⁹ However, Meta is not bound by any policy guidance the Board gives, nor is Meta bound to follow the Board’s advice in the cases it refers to the Board requesting a general policy advisory statement. Thus, when it comes to broader policy, the Board’s input “will be taken into consideration,” and Meta’s only obligation is to publicly respond.⁶⁰

But the Board is not mere *trompe-l’œil* or a façade without substance. Meta made an initial commitment of \$130 million to fund the Board through an independent trust⁶¹ and contributed an additional

(2023) [hereinafter BYLAWS], <https://transparency.fb.com/sr/oversight-board-bylaws-2023> [https://perma.cc/XX38-DA2F].

52. CHARTER, *supra* note 51, art. 1, § 3.

53. *Oversight Board Appoints New Board Members*, OVERSIGHT BD. (May 2022), <https://www.oversightboard.com/news/464999558726685-oversight-board-appoints-new-board-members/> [https://perma.cc/BLM6-394Y].

54. CHARTER, *supra* note 51, art. 2, § 1.

55. *Id.* art. 2, § 2.

56. *Id.*

57. See BYLAWS, *supra* note 51, art. 2, § 1.2.1; Evelyn Douek, “*What Kind of Oversight Board Have You Given Us?*,” U. CHI. L. REV. ONLINE (May 11, 2020), <https://lawreview.blog.uchicago.edu/2020/05/11/fb-oversight-board-edouek/> [https://perma.cc/XZ2H-C7Q9].

58. CHARTER, *supra* note 51, art. 4.

59. *Id.* art. 3, § 4.

60. *Id.*; BYLAWS, *supra* note 51, art. 2, § 2.3.

61. Brent Harris, *An Update on Building a Global Oversight Board*, META (Dec. 12, 2019), <https://about.fb.com/news/2019/12/oversight-board-update/> [https://perma.cc/QV7V-2QA3].

\$150 million to the trust in 2022.⁶² The resumes of the Board’s inaugural members are impressive, including former judges, a Nobel Peace Prize laureate, law professors, political leaders, and journalists,⁶³ and Meta has since asked them to review some of its most high-profile and controversial decisions, starting with the decision of how to handle former President Donald Trump’s account in January 2021.⁶⁴

B. The Objectives

Trying to understand the purpose of the Board only by looking at the ways it is explicitly described in its constituent documents and other public-facing materials does not get one very far. Meta and the Board’s websites both declare that the Board was created to “help [Meta] resolve some of the most difficult questions around freedom of expression online: what to take down, what to leave up and why.”⁶⁵ This is accurate in the most literal sense, but it does not answer the deeper question about what kind of “help” the Board is intended to provide. The fact that there are no engineers amongst the Board’s members suggests, for example, that the Board was not intended to help review the code of Meta’s automated content moderation tools. So what exactly is the deficiency in Meta’s decision-making that the Board was created to address? And how do the choices of the Board’s institutional designers help address those deficiencies?

The Board’s design and context show that substantive decision-making is a relatively minor part of the Board’s intended benefits. Instead, the fundamental goal of the Board is to create legitimacy for Meta⁶⁶ by creating the perception that Meta’s decisions are “justified, appropriate, or otherwise deserving of support for reasons beyond fear of sanctions or mere hope for personal reward.”⁶⁷

62. *Securing Ongoing Funding for the Oversight Board*, OVERSIGHT BD. (July 2022), <https://oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board/> [<https://perma.cc/UE79-LYE7>].

63. See Ben Smith, *Trump Wants Back on Facebook. This Star-Studded Jury Might Let Him*, N.Y. TIMES (Jan. 24, 2021), <https://www.nytimes.com/2021/01/24/business/media/trump-facebook-oversight-board.html> [<https://perma.cc/8L2E-EHXQ>].

64. Nick Clegg, *Referring Former President Trump’s Suspension from Facebook to the Oversight Board*, META (Jan. 21, 2021), <https://about.fb.com/news/2021/01/referring-trump-suspension-to-oversight-board/> [<https://perma.cc/LZ6Y-LRPE>].

65. *The Oversight Board*, META TRANSPARENCY CTR., <https://transparency.fb.com/oversight/> [<https://perma.cc/ND7S-ML8W>]; OVERSIGHT BD., <https://oversightboard.com/> [<https://perma.cc/8FGE-E9MJ>].

66. Douek, *supra* note 44, at 18.

67. Richard H. Fallon, *Legitimacy and the Constitution*, 118 HARV. L. REV. 1787, 1795 (2005); see also Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUST. 283, 307 (2003) (“Legitimacy is the property that a rule or an authority has when others feel obligated to defer voluntarily.”).

For Noah Feldman, the law professor who first proposed the idea of a Facebook Supreme Court to executives at Meta,⁶⁸ the Board did have at least a partially substantive purpose. Feldman originally conceived of the Board as a direct response to the threats to free speech in the social media age and to “the pressure that the platforms face to limit expression in order to satisfy engaged, committed advocacy groups.”⁶⁹ Therefore, Feldman’s goal for the Board was explicitly substantive and values-laden, not merely procedural: the Board should be “an effective counterweight to censorship.”⁷⁰ But the institutional design of the Board was intended to deliver these substantive goals in a way that reduced the decision costs to Meta, and the benefits of “legitimacy” also pervade Feldman’s original memos arguing for the Board’s creation. The Board, he wrote, should become a “durable institution to deliver *principled, reasoned decision-making* that would be widely understood as *legitimate*.”⁷¹ The public-facing, judicial model of the Board was essential to “*capture the legitimacy benefits of decisional independence*.”⁷²

It was primarily these procedural and legitimacy benefits were attractive to Meta. When Zuckerberg was first proposing and explaining his decision to set up the Board, he did refer to the importance of its role in protecting free expression.⁷³ But he “also recogniz[ed] the reality of keeping people safe”⁷⁴ and did not center the Board’s role in pushing back against censorship. Instead, when Zuckerberg set out to sell the Board to the public, he focused on the procedural aspects of its institutional design. He strongly emphasized its intended independence, writing that the Board would “prevent the concentration of too much decision-making within our teams” and “provide assurance that these decisions are made in the best interests of our community and not for commercial reasons.”⁷⁵ Thus, for Zuckerberg, the Board was intended to *outsource* the decision-making, not necessarily to make it *better* or *more speech protective*. The charitable reading of this is that Zuckerberg wanted to benevolently relinquish power to ensure that justice was

68. Klonick, *The Facebook Oversight Board*, *supra* note 44, at 2449.

69. Letter from Noah Feldman to Facebook (Jan. 30, 2018), in OVERSIGHT BD., GLOBAL FEEDBACK AND INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS: APPENDIX, at 101, <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf> [<https://perma.cc/D8PM-5DEC>].

70. *Id.* at 102.

71. Memorandum from Noah Feldman in OVERSIGHT BD., GLOBAL FEEDBACK AND INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS: APPENDIX, at 104, <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf> [<https://perma.cc/D8PM-5DEC>] (emphasis rendered in bold in original).

72. *Id.* at 110 (emphasis rendered in bold in original).

73. See Zuckerberg, *supra* note 49; Mark Zuckerberg Stands for Voice and Free Expression, META (Oct. 17, 2019), <https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/> [<https://perma.cc/P5AT-7RRG>].

74. Zuckerberg, *supra* note 49.

75. *Id.*

not only done but also seen to be done. The more cynical take is that Zuckerberg wanted to throw content moderation's hottest potatoes to other decision-makers and get them off his hands.

Either way, there was interest alignment between Zuckerberg and other stakeholders. There had long been concerns about the astonishing power social media companies wield over global communications, concentrated in the hands of a few. In 2008, Jeffrey Rosen wrote in *The New York Times* about the example of the deputy general counsel of Google, Nicole Wong, who was nicknamed "The Decider" because of her control over what would and would not be allowed on some of the most precious real estate on the Internet.⁷⁶ Rosen suggested that "Wong and her colleagues arguably have more influence over the contours of online expression than anyone else on the planet."⁷⁷ This raised the "increasingly urgent" question, said Rosen: "Can we trust a corporation to be good?"⁷⁸ Nothing much had changed by 2012, when Rebecca MacKinnon wrote about the lack of traditional forms of legitimating constraints on the power of tech platforms and asked, again: "How do we make sure that people with power over our digital lives will not abuse that power?"⁷⁹ As the companies in question only grew in size and importance, these concerns became more pressing, and they reached new heights in the aftermath of the 2016 U.S. presidential election.⁸⁰ The fundamental problem for Meta in that moment was foreshadowed by Professor Tim Wu over a decade and a half before, in the context of one of Meta's competitors: "To love Google, you have to be a little bit of a monarchist . . . [T]hey live and die on trust, and as soon as you lose trust in Google, it's over for them."⁸¹ In the techlash, public trust in platforms hit a nadir, and the platforms were looking for ways to get it back. The Board was the most unique and high-profile attempt, purporting to finally provide an answer for how to build back confidence in Meta's corporate monarchy: separation of powers.

To be clear, the interest alignment only went so far. Lest one think that the Board was a genuine answer to critics' long-standing fears of Meta's unchecked power over speech, the Board's limited remit, or "jurisdiction," is a clear reminder that Meta's goal in setting up the Board did not match its rhetoric. As Nick Clegg, Meta's President of Global Affairs, wrote when welcoming the Board's first members: "[W]e have created and empowered a new group to exercise independent judgment

76. Jeffrey Rosen, *Google's Gatekeepers*, N.Y. TIMES MAG. (Nov. 28, 2008), <https://www.nytimes.com/2008/11/30/magazine/30google-t.html> [<https://perma.cc/299Z-WEH7>].

77. *Id.*

78. *Id.*

79. REBECCA MACKINNON, CONSENT OF THE NETWORKED: THE WORLDWIDE STRUGGLE FOR INTERNET FREEDOM xxii (2012).

80. *See, e.g.*, Smith, *supra* note 47.

81. Rosen, *supra* note 76.

over some of the most difficult and significant content decisions.”⁸² “Some” was right. The areas of decision-making over which the Board actually had oversight were, and remain, extremely limited.⁸³ For example, the Board does not have the power to review decisions Meta makes about accounts, advertising, events, groups, amplification, or the platforms’ features and affordances.⁸⁴ When the Board was first created, it did not even have the power to review cases in which Meta decided to leave content *up* as opposed to when it took content down.⁸⁵ And while this has since been remedied, there are few signs the Board’s “jurisdiction” is set to expand any further.

Thus, in many ways, Meta’s creation of the Board resembles authoritarian governments’ use of courts.⁸⁶ Authoritarian rulers maintain courts because courts can provide a patina of legitimacy,⁸⁷ they can provide a measure of predictability which can be useful for attracting commercial investment in particular,⁸⁸ and they allow for the deflection of controversy away from the ruling regime.⁸⁹ But in order to get these benefits without too great a cost, such rulers often confine the scope of courts’ jurisdiction such that “[a] relatively independent judiciary may be preserved but simply excluded from domains significant to the authoritarian regime.”⁹⁰ As a result, the actual trust garnered by such institution tends to be limited,⁹¹ and the mainstream initial reception to

82. Clegg, *supra* note 50.

83. I wrote about this far too often during the process leading up to the Board’s creation. See, e.g., Douek, *supra* note 44, at 39–46; Douek, *supra* note 57; Evelyn Douek, *The Facebook Oversight Board Should Review Trump’s Suspension*, LAWFARE (Jan. 11, 2021, 11:15 AM), <https://www.lawfareblog.com/facebook-oversight-board-should-review-trumps-suspension> [<https://perma.cc/AJT6-QBP5>].

84. BYLAWS, *supra* note 51, art. 3, § 1.1.2.

85. Douek, *supra* note 57; Guy Rosen, *Users Can Now Appeal Content Left Up on Facebook or Instagram to the Oversight Board*, META (Apr. 13, 2021), <https://about.fb.com/news/2021/04/users-can-now-appeal-content-left-up-on-facebook-or-instagram-to-the-oversight-board/> [<https://perma.cc/F62K-HMC4>].

86. Douek, *supra* note 44, at 9–10.

87. See Tamir Moustafa, *Law and Courts in Authoritarian Regimes*, 10 ANN. REV. L. & SOC. SCI. 281, 286 (2014).

88. Tamir Moustafa, *Law and Resistance in Authoritarian States: The Judicialization of Politics in Egypt*, in RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES 132, 132, 138 (Tamir Moustafa & Tom Ginsburg eds., Cambridge Univ. Press 2008).

89. Tamir Moustafa & Tom Ginsburg, *Introduction: The Functions of Courts in Authoritarian Politics*, in RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES 1, 10 (Tamir Moustafa & Tom Ginsburg eds., Cambridge Univ. Press 2008); see also Mark A. Graber, *The Nonmajoritarian Difficulty: Legislative Deference to the Judiciary*, STUD. AM. POL. DEV., Spring 1993, at 35, 43.

90. Martin Shapiro, *Courts in Authoritarian Regimes*, in RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES 326, 331 (Tamir Moustafa & Tom Ginsburg eds., Cambridge Univ. Press 2008).

91. See Mark Tushnet, *Authoritarian Constitutionalism*, 100 CORNELL L. REV. 391, 422–25 (2015).

the Board reflected this skepticism.⁹² But in order to get any of the legitimacy and trust dividends that Meta hoped for from the Board, it needed to meaningfully relinquish some power. The balance that Meta struck in the Board's design was fairly robust independence mechanisms within an extremely limited jurisdiction.

There are of course many different kinds of legitimacy and different theories about how to establish it, and the Board's model was left unspecified and undertheorized.⁹³ One mechanism through which the Board might further create legitimacy was by increasing Meta's accountability and transparency,⁹⁴ in the sense that it would require Meta to provide at least a bare minimum amount of information about, and justification for, its decisions.⁹⁵ This is a thin definition of accountability, but even so it is one that is sorely lacking across the content moderation industry.

Some also hoped the Board would build legitimacy by mitigating long-standing concerns over the democratic deficits of content moderation.⁹⁶ This is similar to the goal of outsourcing power away from Meta as the sole decision-maker but is more specific. It is not only that such significant power should not lie solely in the hands of a single

92. See, e.g., Vaidhyathan, *supra* note 31; Sara Fischer, *Facebook Critics Take on Its Oversight Board*, AXIOS (Sept. 25, 2020), <https://www.axios.com/2020/09/25/facebook-critics-take-on-its-oversight-board> [<https://perma.cc/D3JB-T784>].

93. See generally Brenda Dvoskin, *Expertise and Participation in the Facebook Oversight Board: From Reason to Will*, TELECOMMS. POL'Y, June 2023, at 1; Blayne Haggart & Clara Iglesias Keller, *Democratic Legitimacy in Global Platform Governance*, TELECOMMS. POL'Y, June 2021, at 7–9.

94. See, e.g., Oreste Pollicino & Giovanni De Gregorio, *Shedding Light on the Darkness of Content Moderation: The First Decisions of the Facebook Oversight Board*, VERFASSUNGSBLOG (2021), <https://verfassungsblog.de/fob-constitutionalism/> [<https://perma.cc/3VY4-JYL2>] (hoping the Board will “increase transparency in content moderation, while making Facebook more accountable for its choices”); David Wong & Luciano Floridi, *Meta's Oversight Board: A Review and Critical Assessment*, 33 MINDS & MACHS. 261, 266 (2023) (observing that the Board's “significant strengths [include] its ability to enhance the transparency of content moderation decisions and processes”).

95. This definition of accountability draws on, for example, Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 404, 438 (2006); Jody Freeman, *The Private Role in the Public Governance*, 75 N.Y.U. L. REV. 543, 664 (2000); Jerry L. Mashaw, *Accountability and Institutional Design: Some Thoughts on the Grammar of Governance*, in PUBLIC ACCOUNTABILITY: DESIGNS, DILEMMAS, AND EXPERIENCES 115, 117 (Michael W. Dowdle ed., Cambridge Univ. Press 2006).

96. Klonick, *The Facebook Oversight Board*, *supra* note 44, at 2499 (stating that the Board “signifies a step towards empowering users by involving them in private platform governance”); Edward Lee, *Virtual Governments*, 27 UCLA J.L. & TECH. 1, 5 (2022) (arguing that platforms “operate in the shadow of a democratic deficit”); Barrie Sander, *Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation*, 43 FORDHAM INT'L. L.J. 939, 991 (2020) (stating that “Facebook and other major platforms could go much further in ensuring more structured multi-stakeholder participation in the development and revision of their content moderation rules”).

corporate entity, but that it should be in some meaningful way informed by broader public participation.⁹⁷

But most fundamentally, the Board’s design is centered around creating legitimacy by replicating a particular form of speech governance that is “afforded a degree of presumptive legitimacy”⁹⁸ — that is, by acting like a court. This was, of course, the initial framing that Feldman and Zuckerberg suggested for the Board, and it has had remarkable staying power, “successfully shift[ing] the discourse around the [Board] into a register that invited comparisons between Facebook and the legal systems of democratically founded states, rather than to classic forms of industry-led informal regulation and/or lobbying.”⁹⁹ The most significant feature of judicial decision-making that the Board’s institutional design replicates is the obligation to provide *public rationales* for its decisions.¹⁰⁰ As Feldman argued in his original proposal, “[t]he key to making this private-courts approach work is recognizing that there is no magic-bullet solution to balancing competing values The advantage enjoyed by real-life constitutional courts is that they openly address difficult cases, and so derive credit and legitimacy from being principled.”¹⁰¹ The duty to publish reasons might seem self-evident to lawyers accustomed to this practice¹⁰² but was not inevitable. The Board could have been set up to provide independent but *in camera* review of Meta’s rules — many platforms have trust and safety advisory boards or similar bodies that play this kind of role.¹⁰³ In departing from this model, the Board’s designers made a conscious choice to require public reasons, along with the directive that such decisions will have precedential effect.¹⁰⁴ This gives the Board an important

97. See *infra* Section III.A.5 (discussing the ambiguity in what exactly that public participation should look like).

98. David Landau & Rosalind Dixon, *Abusive Judicial Review: Courts Against Democracy*, 53 U.C. DAVIS L. REV. 1313, 1335 (2020).

99. Michael Veale, Kira Matus & Robert Gorwa, *AI and Global Governance: Modalities, Rationales, Tensions*, 19 ANN. REV. L. SOC. SCI. 225, 268 (2023) (citations omitted); see also Kadri, *supra* note 44, at 171–74; JOSH COWLS, PHILIPP DARIUS, DOMINIQUE SANTISTEVAN & MORITZ SCHRAMM, CONSTITUTIONAL METAPHORS: FACEBOOK’S “SUPREME COURT” AND THE LEGITIMATION OF PLATFORM GOVERNANCE 15 (2022), <https://doi.org/10.1177/14614448221085559> [<https://perma.cc/PWA6-7Q5H>].

100. See BYLAWS, *supra* note 51, art. I, §§ 3.1.7, 3.2 (2022).

101. Feldman, *supra* note 69, at 102.

102. Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 633 (1995) (“The practice of providing reasons for decisions has long been considered an essential aspect of legal culture.”).

103. See, e.g., *Introducing the Spotify Safety Advisory Council*, SPOTIFY (June 13, 2022), <https://newsroom.spotify.com/2022-06-13/introducing-the-spotify-safety-advisory-council/> [<https://perma.cc/367P-VTPX>]; *Safety Advisory Council*, TWITCH (May 2020), <https://safety.twitch.tv/s/article/Safety-Advisory-Council> [<https://perma.cc/9BAK-NHRZ>]; Trust & Safety Pro. Ass’n, *Approaches to Trust & Safety*, TRUST & SAFETY PRO. ASS’N (May 24, 2022), <https://www.tspa.org/curriculum/ts-fundamentals/industry-overview/ts-approaches/> [<https://perma.cc/B8ZL-T4MP>].

104. CHARTER, *supra* note 51, art. 2, § 2.

explanatory role, and reflects a hope that the Board will provide substantial and coherent public justifications for Meta’s otherwise opaque rules and decision-making.¹⁰⁵

The Board itself has, in appearance at least, leaned into this conception of its role.¹⁰⁶ The Board’s decisions clearly ape the norms and forms of a judicial institution. The Board has stuck to an IRAC (Issue, Rule, Analysis, Conclusion) template in its decisions, familiar to every law student, and applied it mechanically and faithfully in every case. It makes “findings,” cites “sources” for its authority, summarizes parties’ “submissions,” and invokes the decision-making rubric of IHRL.¹⁰⁷ It “overturns” or “upholds” Meta’s original enforcement actions and has started releasing “summary decisions.”¹⁰⁸ In all these ways, the Board is seeking to invoke the legitimacy afforded to legal decision-making.

Thus, the Board has many subsidiary goals: to make content moderation more speech protective, independent, transparent, and accountable, perhaps even more democratic, and to provide public reasoning. But at its root, these goals are all in service of the Board’s overarching purpose: to establish legitimacy as a source of authority for important and difficult disputes about online speech rules.¹⁰⁹

III. REVIEWING THE BOARD’S PERFORMANCE

According to the Board, it is doing a very good job indeed of achieving its goals. The Board’s inaugural members have completed their first three-year term, and every member that wanted to extend their

105. Douek, *supra* note 44, at 66–67.

106. Kadri, *supra* note 44, at 174 (giving examples of how Board members have embraced the judicial analogy).

107. See, e.g., *Trump Suspension Case*, *supra* note 1.

108. *Oversight Board Publishes First Summary Decisions*, OVERSIGHT BD. (June 2023), <https://oversightboard.com/news/1882578038790997-oversight-board-publishes-first-summary-decisions/> [https://perma.cc/23KA-49PV].

109. See Douek, *supra* note 44, at 18 (“Legitimacy is central to the FOB experiment.”); Klonick, *The Facebook Oversight Board*, *supra* note 44, at 2427 (“Facebook’s creation of the Oversight Board is an investment in building user trust”); Monroe E. Price & Joshua M. Price, *Building Legitimacy in the Absence of the State: Reflections on the Facebook Oversight Board*, 17 INT’L J. COMM’N 3315, 3319 (2023) (“The aspiration is that the insertion of this new element — the Oversight Board — becomes a productive and effective attribute of legitimacy.”); Chinmayi Arun, *Facebook’s Faces*, 135 HARV. L. REV. F. 236, 240 (2022) (“[T]he Oversight Board was created to bring it more legitimacy.”); Gilad Abiri & Sebastian Guidi, *From a Network to a Dilemma: The Legitimacy of Social Media*, 26 STAN. TECH. L. REV. 92, 115 (2022) (“The idea, in short, was that the Board would enhance Meta’s legitimacy”); Matthias C. Kettemann & Wolfgang Schulz, *How to Integrate Public Values into Private Orders*, in PLATFORMS//DEMOCRACY: PERSPECTIVES ON PLATFORM POWER, PUBLIC VALUES AND THE POTENTIAL OF SOCIAL MEDIA COUNCILS 6, 7 (Matthias C. Kettemann & Wolfgang Schulz eds., 2023) (arguing social media councils like the Board “can provide more legitimacy to the rules and algorithmic practices of platforms”); Haggart & Keller, *supra* note 93 (discussing the different ways the Board could establish legitimacy).

tenure was renewed for a second.¹¹⁰ The Board's Trustees, who oversee the process for approving a Board member's reappointment,¹¹¹ thus appear happy with the Board's overall performance. The Board's own assessment is similarly positive. In the Board's eyes, "early successes in holding Meta accountable demonstrate the Board's viability and provide a self-regulatory framework for extending and improving our operations in the future."¹¹² Through its work, "the Board has demonstrated" the importance of its role.¹¹³

This Part reviews the Board's performance so far and whether these claims are justified. First, I look at the Board's successes, which are not insubstantial. At a minimum, the Board has succeeded at being a functional institution in the most literal sense, by hearing cases, making decisions, and issuing recommendations to Meta, and it has succeeded in avoiding the failure scenario most commonly predicted for it: it has not been captured by Meta and has asserted its independence in significant ways. The Board has also succeeded in some of the more ambitious goals people had for it — it has made Meta's content moderation more accountable, transparent, and participatory and has garnered public attention and respect for doing so. But the Board's claims of success need to be qualified. Section III.B turns to the Board's missed opportunities and the ways in which it has failed to live up to its potential — its lack of attention to developing a normative framework for thinking about private content moderation, its measurement of impact based on metrics that are poor proxies for material benefit, and its persistent avoidance in the hardest cases.

Overall, the story of this Part is ambiguous. The Board is perhaps much more successful than many might have anticipated. But even if it did not fail in expected ways — being captured by Meta, becoming forgotten by content moderation stakeholders, having its decisions consistently overruled — it has failed to live up to its potential in other ways that have not received enough attention.

A. The Board's Success Stories

When the Board was established, no one knew what to expect. It was an unprecedented experiment in governance for a private

110. *Updates on Oversight Board Membership*, OVERSIGHT BD. (Apr. 2023), <https://oversightboard.com/news/771690787717546-updates-on-oversight-board-membership/> [<https://perma.cc/6AD8-G2EH>].

111. CHARTER, *supra* note 51, art. 1, § 3.

112. OVERSIGHT BD., 2021 ANNUAL REPORT 5 (2021) [hereinafter 2021 ANNUAL REPORT].

113. @OversightBoard, X (FORMERLY KNOWN AS TWITTER) (Nov. 4, 2022, 8:50 AM), <https://twitter.com/OversightBoard/status/1588513975188484099> [<https://perma.cc/WRS8-WS6H>].

company — or, put another way, “a bit odd.”¹¹⁴ It was therefore impossible to predict what it might achieve. Would the Board become a model for the industry? Or would it fall apart and join the graveyard of other all-but-forgotten experiments in social media governance, like the time Facebook tried letting users vote on its rules?¹¹⁵ Only a few years in, a final verdict is still premature. But so far, it seems the Board is, in many ways, working. This Section reviews some of the success stories of the Board. But each of these stories is also one of limitation — every achievement of the Board also reveals the constraints on what it can do or has done.

1. Fulfilling its Basic Functions

To start with, the Board has succeeded at performing its role in the most literal sense.¹¹⁶ The functional purpose of the Board is to hear cases, issue opinions, and give Meta recommendations. Over the past three years, the Board has done exactly that. This is more of a minimum condition for success than a dramatic achievement, but with any experimental institution, there is no guarantee of even this form of success.

One way of reviewing an institution’s output is by looking at some headline statistics,¹¹⁷ and indeed the executive summary of the Board’s Annual Reports tend to focus on this kind of data in summarizing the Board’s performance. For example, a reader opening the Board’s first Annual Report will be informed that in its first year, the Board received over 1.1 million appeals, published twenty decisions, overturned Meta’s decisions seventy percent of the time, made eighty-six policy recommendations, which Meta took steps towards implementing two-thirds of the time,¹¹⁸ and so on.¹¹⁹

Numbers like these are clearly intended to tell a story: there is high demand for the Board (over a million appeals!), the Board is being tough on Meta (overturning so many of its decisions!), the number of recommendations it makes far outstrips the bare numbers of decisions it makes (eighty-six recommendations versus twenty decisions!), and its power is evident in how often Meta has committed to reform as a result of the Board’s work (two-thirds of the time!). The exact numbers

114. David Fontana & David Schleicher, *The Basketball Court 4* (GWU Legal Studs. Rsch. Paper No. 2023-38, 2023).

115. Adi Robertson, *Facebook Used to Be a Democracy — but Nobody Voted*, VERGE (Apr. 5, 2018), <https://www.theverge.com/2018/4/5/17176834/mark-zuckerberg-facebook-democracy-governance-vote-failure> [<https://perma.cc/4S7N-YC26>].

116. This Article incorporates data from the Board’s decisions up to June 30, 2023.

117. See, e.g., *The Statistics*, 136 HARV. L. REV. 500 (2022) (providing statistics on the most recent Supreme Court term).

118. 2021 ANNUAL REPORT, *supra* note 112, at 6–7 (providing statistics for 2021).

119. The Board’s subsequent Annual Reports have continued this practice. See, e.g., 2022 ANNUAL REPORT, *supra* note 21, at 8–9 (providing statistics for 2022).

have changed in subsequent years, but the overall gist, and the story they are intended to tell, have remained the same.

It is important to put this story in context, however. Pushing a little on these numbers reveals a more complicated picture than the one that the Board wants to portray. Let's take a few illustrative examples.

First, the Board often points to the fact that many users have lodged appeals to the Board to indicate “enormous pent-up demand” for additional review mechanisms.¹²⁰ The Board's annual report for 2022 proudly proclaims that “[o]n average, the Board received a case every 24 seconds”¹²¹ The Board highlights this figure to suggest that there is “ongoing demand from users to appeal Meta's content moderation decisions to an independent body.”¹²² This figure of course does show that there are some users who remain dissatisfied with Meta's rules and internal appeals processes, but by itself this number actually tells us little about the importance of the Board or the general public demand for such an institution. In reality, in context the number is actually astonishingly *small*. The Board received an average of 3,537 appeals per day in 2022¹²³ while Meta performed some sort of content moderation on 35,608,178 *pieces of content a day* in the last quarter of that period (a figure that doesn't include the number of times Meta decided to leave content *up* after someone flagged it for review, which could also be appealed to the Board) and received 140,598 internal appeals every day.¹²⁴ That is, users only appealed what would be *significantly* less than 0.0001% of Meta's decisions to the Board, and only 2.5% of users who appealed decisions to Meta went the additional step of appealing to the Board. The level of demand for the Board thus looks very different when viewed in relative, rather than absolute, terms.

Even this figure lacks context. The Board's citation of raw numbers does not indicate the nature or quality of these appeals. Content moderation experience suggests that these appeals would be a very noisy indicator of problems in Meta's enforcement systems.¹²⁵ Meta has previously stated, for example, that when it gave users the opportunity to

120. 2021 ANNUAL REPORT, *supra* note 112, at 4.

121. 2022 ANNUAL REPORT, *supra* note 21, at 30 (emphasis in original omitted).

122. *Id.* at 32.

123. *Id.* at 30.

124. *Quarterly Community Standards Enforcement Report*, META, <https://transparency.fb.com/reports/community-standards-enforcement/> [<https://perma.cc/HPT9-57LJ>] (using the linked CSV data, computed by aggregating totals for “Content Actioned” and “Content Appealed” in Q42022, then dividing the total by ninety-two, the number of days in October, November, and December).

125. Brian Fishman, *Dual-Use Regulation: Managing Hate and Terrorism Online Before and After Section 230 Reform*, BROOKINGS (Mar. 14, 2023), <https://www.brookings.edu/articles/dual-use-regulation-managing-hate-and-terrorism-online-before-and-after-section-230-reform/> [<https://perma.cc/T5Y2-C9JA>] (“One of the most dispiriting early lessons I learned at Meta was that user reports, paradoxically, are both critically important and wildly unreliable.”).

provide more information to support their appeals, only two percent of the information people provided was useful.¹²⁶ Therefore, the simple fact that there are a lot of appeals does not in itself make the case that the Board is necessary to satisfy pent up and sincere demand for an additional appeals mechanism.

The Board is struggling to meet the demand that there is, however. This is a predictable problem — the scale of content moderation is a fundamental problem that any system of content moderation governance must reckon with and suggests the limits of an approach based around *ex post*, individual appeals.¹²⁷ But even accepting that the Board can only decide a limited fraction of the appeals it receives, the Board is underdelivering. We can see this if we look at the unimpressive statistics about how the Board has met the overwhelming demand for it. In total, the 20+ member Board issued forty-five decisions in the first *three years* of operation.¹²⁸ By comparison, in the 2021 term alone, the nine-member Supreme Court issued full opinions in sixty cases.¹²⁹

The paucity of issued opinions is not the consequence of a paucity of funding. As Helfer and Land note, out of all the international human rights tribunals, only the European Court of Human Rights (“ECHR”) has a budget that is anywhere close to the size of the Board’s, yet the ECHR reviews more cases by several orders of magnitude per year than the Board (36,000 vs. twenty in 2021).¹³⁰ Board decisions are also significantly shorter and generally less detailed than a typical appellate court decision. For people earning six-figure salaries,¹³¹ the Board’s output is incredibly low (no doubt in part because Board positions are considered part-time roles). Worse yet, these relatively few and short decisions often arrive after the deadlines set for them in the Board’s own bylaws.¹³² In a sign of the quasi-make-believe nature of the Board, the fact that essentially every decision is now issued in breach of the procedures laid out in the Board’s notionally constituent documents appears to trouble no one involved in the project. The Board simply notes

126. *Oversight Board Selects Case Regarding a Post Depicting Indigenous Artwork and Discussing Residential Schools*, META TRANSPARENCY CTR. (June 12, 2023), <https://transparency.fb.com/oversight/oversight-board-cases/indigenous-artwork-residential-schools> [<https://perma.cc/PAS9-KPNF>].

127. Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 548–56 (2022).

128. As of June 30, 2023.

129. *The Statistics*, *supra* note 117, at 500.

130. Laurence R. Helfer & Molly K. Land, *The Meta Oversight Board’s Human Rights Future*, 44 CARDOZO L. REV. 2233, 2254–55 (2023).

131. Allana Akhtar, *Facebook is Reportedly Paying Its Oversight Board 6-Figure Salaries Only for It to Tell the Company to Solve Its Own Problems*, BUS. INSIDER (May 6, 2021), <https://www.businessinsider.com/facebook-oversight-board-members-get-six-figure-salaries-report-2021-2> [<https://perma.cc/6W9M-2UPF>].

132. Evelyn Douek & Tia Sewell, *Meta’s Oversight Board Often Turns in Its Homework Late. Does It Matter?*, LAWFARE (July 15, 2022), <https://www.lawfareblog.com/metas-oversight-board-often-turns-its-homework-late-does-it-matter> [<https://perma.cc/6W9M-2UPF>].

the reasons why decisions are late (which include members taking leave and scheduling challenges) in footnotes in the Board's transparency reports.¹³³

While it is true that much of the impact of the Board will come through its broader recommendations rather than its individual decisions, the Board was purposefully set up in an adjudicative model in which deciding individual cases is a core part of its role. As one tech reporter put it, "It's true that the board's actions go beyond the three cases it decided [that quarter] At the same time, I can't be the only person to feel like the board is slacking."¹³⁴ So even as the Board emphasizes overwhelming demand for its work, its output suggests that it is struggling to meet that demand. Meta has implicitly acknowledged the problem of low and slow work product from the Board, noting its hopes and expectations that the Board will "significantly increase its output and impact"¹³⁵ by starting to take expedited reviews and issue summary decisions.

There would appear to be another simple way to at least increase the baseline from the currently incredibly low output: appoint more Board members. The Board was originally intended to consist of forty members,¹³⁶ but in its three years, the Board has only appointed new Board members three times.¹³⁷ In April 2023, it announced it would not be appointing many more: "While we originally expected the Board to reach 40 Members, three years of operations has shown us that, in practice, the optimal number of Members allowing for timely, regular, and effective deliberation and decision-making, is 26."¹³⁸ This very specific figure (twenty-six, not one more or less!) is offered without explanation, and no mention is made of the countervailing considerations, like the fact that the Board appears to be unable to manage its workload, or the impacts on the ability to secure representation from certain geographic regions. Again, the figures given in the Charter and Bylaws for the composition of the Board appear to be seen by the Board as a suggestion only. Importantly, perhaps tellingly, at least one effect of the

133. See, e.g., OVERSIGHT BD., Q1 2023 QUARTERLY TRANSPARENCY REPORT 13 n.8 (2023) [hereinafter Q1 2023 QUARTERLY TRANSPARENCY REPORT].

134. Casey Newton, *The Oversight Board Spins Its Wheels*, PLATFORMER (Oct. 20, 2022), <https://www.platformer.news/p/the-oversight-board-spins-its-wheels> [https://perma.cc/5DLV-YAM7].

135. META, META Q1 2023 QUARTERLY UPDATE ON THE OVERSIGHT BOARD 8 (2023).

136. CHARTER, *supra* note 51, art. 1, § 1; BYLAWS, *supra* note 51, art. 1, § 1.4.

137. See *Announcing a New Board Member, and Our Next Case*, OVERSIGHT BD. (Apr. 2021), <https://www.oversightboard.com/news/1446487489030285-announcing-a-new-board-member-and-our-next-case/> [https://perma.cc/NP67-4WVL]; *Oversight Board Appoints New Board Members*, *supra* note 53. Although Meta was involved in the selection of the original members, the Board now has sole control over its membership.

138. *Updates on Oversight Board Membership*, OVERSIGHT BD. (Apr. 2023), <https://oversightboard.com/news/771690787717546-updates-on-oversight-board-membership/> [https://perma.cc/P44M-DFRF].

decision not to appoint more members is that the power of the original Board members is not diluted by new membership.

Another metric that is somewhat misleading is the Board's statistics about how often it "overturns" Meta. The way this figure is presented suggests that the Board is more confrontational with Meta than it really is. When the Board announces that it has "overturned" Meta, the suggestion is that the Board and Meta disagreed on the outcome of a case, but the Board's decision prevailed. However, in many instances in which the Board says it has "overturned" Meta, the Board and Meta do not disagree about the right outcome. This is because frequently, once the Board selects a case, Meta reviews the content in question and finds a mistake in its original enforcement action.¹³⁹ It therefore agrees with the Board that the decision was wrong and reverses its decision. In these cases, though, the Board will still announce it has "overturn[ed]" Meta's original decision.¹⁴⁰ To be clear: it is not that the Board is not serving a valuable function in these cases — the Board is identifying errors that would have remained uncorrected if it weren't for the Board selecting the case for review, and the Board is highlighting the existence between Meta's policies on paper and how they are enforced. But identifying mistakes is a different role from forcing Meta to change its normative judgments. The former is more akin to auditing Meta's systems rather than helping Meta "resolve some of the most difficult questions around freedom of expression online."¹⁴¹ For the Board, focusing on a statistic about "overturn" rates defined so broadly makes it look tougher on Meta and helps bolster its public narrative of vigorous and independent oversight.¹⁴²

What these few examples show is that there is no simple way of measuring the Board's work, and the statistics the Board gives about its performance need to be viewed critically and in context. These figures are selected to tell a particular story and to show the Board in the most positive light. There is nothing inherently wrong or unexpected about the Board or Meta trying to make themselves look as good as possible, but it requires approaching their statements with caution and with an understanding of the picture the handpicked figures are intended to paint. And it underscores why an in-depth study of the Board's work is necessary before it can be properly evaluated.

139. 2022 ANNUAL REPORT, *supra* note 21, at 34 ("Meta deemed its original decision in 32 out of 50 cases shortlisted in 2022 (64%) to have been incorrect . . .").

140. *See, e.g.*, Mention of the Taliban in News Reporting Case, 2022-005-FB-UA (Oversight Bd. Sept. 15, 2022), <https://www.oversightboard.com/decision/FB-U2HHA647> [<https://perma.cc/2H5T-L5WL>]; Colombian Police Cartoon Case, 2022-004-FB-UA (Oversight Bd. Sept. 15, 2022), <https://www.oversightboard.com/decision/FB-I964KKM6> [<https://perma.cc/VC7C-FCPP>].

141. *The Oversight Board*, *supra* note 65.

142. *Id.* & Floridi, *supra* note 94, at 8 (suggesting that a "notable strength of the [Board] is its assertiveness, manifested in its willingness to overrule Meta").

2. The Board's Assertions of Independence

The Board has been successful in defining its identity as an independent entity and not merely one acting at the behest of its creator. A core feature of the Board — indeed, perhaps its most novel aspect as an experiment in corporate governance — is that it would exercise “independent judgment”¹⁴³ and not be beholden to Meta. But the self-regulatory nature of the Board naturally led many people, particularly at first, to express concerns about whether it could be sufficiently independent, given it was created and funded by the entity it is intended to oversee.¹⁴⁴ However, the first few years of the Board's existence have largely dispelled critiques of the Board model on the grounds that it would inevitably be captured and compromised by Meta. The Board has asserted a significant degree of independence in its first few years, including in ways that depart from the vision that Meta originally had for the Board and its role.

To be clear, Meta did not give the Board completely free rein. Indeed, the remit of the Board ensures it cannot function as a comprehensive oversight body, by only allowing it to review a small portion of Meta's decision-making. Nevertheless, Meta did give the Board substantial independence by putting the Board's funding into an irrevocable trust¹⁴⁵ and placing the power to remove Board members exclusively in the hands of independent trustees.¹⁴⁶ This of course was not selfless. Meta could not get the legitimacy dividends it wanted without such independence. For this reason, within its limited remit, the mechanisms protecting the Board's independence are robust. This has protected the Board's decisional independence but has also given the Board the opportunity to depart from the precise plans Meta had for it.

Perhaps most consequentially, the Board's recommendations have been broader than originally envisioned. Meta has remarked that “[t]he size and scope of the board's recommendations go beyond the policy guidance that we first anticipated when we set up the board, and several require multi-month or multi-year investments.”¹⁴⁷ The recommendations have, amongst other things, targeted the systems underlying Meta's enforcement mechanisms (e.g., to “[e]nsure users can appeal

143. *The Oversight Board*, *supra* note 65.

144. See Alex Kantrowitz, *Should We Trust the Facebook Oversight Board?*, BIG TECH. (Apr. 15, 2021), <https://www.bigtechnology.com/p/should-we-trust-the-facebook-oversight> [<https://perma.cc/8W5G-432Q>] (“Critics of the board say it's not actually independent”); FLYNN COLEMAN, BRANDIE NONNECKE & ELIZABETH M. RENIERIS, *THE PROMISE AND PITFALLS OF THE FACEBOOK OVERSIGHT BOARD 1* (2021), https://carcenter.hks.harvard.edu/files/cchr/files/facebook_oversight_board.pdf [<https://perma.cc/9TRG-ECSB>] (“[I]t cannot escape the inherent conflict that all members are on the payroll of the conglomerate.”).

145. OVERSIGHT BOARD TRUST § 1.4 (2019), <https://about.fb.com/wp-content/uploads/2019/12/Trust-Agreement.pdf> [<https://perma.cc/HLY4-C9MK>].

146. See BYLAWS, *supra* note 51, art. 1, § 1.2.2.

147. FACEBOOK, Q1 2021 QUARTERLY UPDATE ON THE OVERSIGHT BOARD 6 (2021).

decisions taken by automated systems to human review”¹⁴⁸ and “[i]nform users when automation is used to take enforcement action against their content”¹⁴⁹), asked Meta to submit to further independent scrutiny and assessment (e.g., to “[e]ngage an independent entity not associated with either side of the Israeli-Palestinian conflict to conduct a thorough examination”¹⁵⁰), and demanded further transparency around sensitive matters (e.g., to “[f]ormalize a transparent process on how it receives and responds to all government requests for content removal”¹⁵¹). Such recommendations require more dramatic restructuring and resources to implement than the simple policy recommendations or rule changes it seems Meta had intended.

The Board has also asserted the power to interpret its own founding documents and the procedures they set out. It has developed what might be described as a mootness doctrine that requires Meta to submit to review (and thus have to answer questions and receive recommendations) even in cases where Meta admits a mistake in the individual case in question.¹⁵² In one of its first cases, Meta argued that the Board should decline to take the case because it had already restored the post in question, agreeing that it had made an error in originally removing the content.¹⁵³ But the Board rejected that argument, saying Meta had misinterpreted the Board’s Charter and that “[f]or Facebook to correct errors the Board brings to its attention and thereby exclude cases from review would integrate the Board inappropriately to Facebook’s internal process and undermine the Board’s independence.”¹⁵⁴ The Board’s capacity to reject Meta’s interpretation of the Charter that Meta itself wrote illustrates a kind of independence and aggressiveness that allows the Board to, within limits, define its role for itself.

The Board’s centering of IHRL as the authoritative body of norms governing its (and therefore Meta’s) decision-making is also an expression of its independence. The Board’s Charter only requires the Board to “pay particular attention” to “human rights norms protecting free expression,”¹⁵⁵ and does not reference IHRL directly at all. The Bylaws

148. Breast Cancer Symptoms and Nudity Case, 2020-004-IG-UA (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/IG-7THR3S11> [<https://perma.cc/LEU7-H9HH>].

149. *Id.*

150. Shared Al Jazeera Post Case, 2021-009-FB-UA (Oversight Bd. Sept. 14, 2021), <https://www.oversightboard.com/decision/FB-P93JPX02> [<https://perma.cc/WY6F-NK4B>].

151. *Id.*

152. *See, e.g.*, Breast Cancer Symptoms and Nudity Case, 2020-004-IG-UA (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/IG-7THR3S11> [<https://perma.cc/LEU7-H9HH>].

153. *Id.*

154. *Id.* (emphasis omitted).

155. CHARTER, *supra* note 51, art. 2, § 2.

make reference to “international human rights standards”¹⁵⁶ but do not require the Board to apply IHRL directly. Nevertheless, the Board has elevated IHRL as its primary source of authority, citing it in every decision,¹⁵⁷ and even overturning one of Meta’s decisions because it found that the company’s own community standards did not comply with IHRL.¹⁵⁸ That is, the Board overruled the authority under which it was created based on an external body of norms.

Finally, the Board has engaged in a perhaps unanticipated amount of what might be described as “Oversight Board Overspeech,” following Josh Chafetz’s description of Congress’s use of its oversight mechanisms to communicate with the broader public as “congressional overspeech.”¹⁵⁹ The Board has actively courted public attention. Its Co-Chairs wrote a New York Times op-ed announcing its creation.¹⁶⁰ It has an Instagram account,¹⁶¹ a LinkedIn account,¹⁶² and an active Twitter account that retweets positive coverage of the Board¹⁶³ and sends out cautionary warnings to Meta in high-stakes moments such as the invasion of Ukraine.¹⁶⁴ The Board received press coverage for issuing “a strong reprimand” to Meta for giving misleading answers to the Board’s questions in one case,¹⁶⁵ and it announced a meeting with a Facebook whistleblower to get more information (although never releasing information about what was said at the meeting).¹⁶⁶ Its members

156. BYLAWS, *supra* note 51, art. 1, § 1.4.4; *see also id.* art. 2, § 2.3.2 (referencing “human rights principles”).

157. *See* JOAN BARATA, THE DECISIONS OF THE OVERSIGHT BOARD FROM THE PERSPECTIVE OF INTERNATIONAL HUMAN RIGHTS LAW 7 (2022), <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2022/10/The-Decisions-of-the-OSB-from-the-Perspective-of-Intl-Human-Rights-Law-Joan-Barata-pdf> [<https://perma.cc/SKA5-VHCM>].

158. Pro-Navalny Protests in Russia Case, 2021-004-FB-UA (Oversight Bd. May 26, 2021), <https://www.oversightboard.com/decision/FB-6YHRXHZR> [<https://perma.cc/WH9R-DNLE>].

159. Josh Chafetz, *Congressional Overspeech*, 89 FORDHAM L. REV. 529, 536 (2020).

160. Catalina Botero-Marino, Jamal Greene, Michael W. McConnell & Helle Thorning-Schmidt, *We Are a New Board Overseeing Facebook. Here’s What We’ll Decide*, N.Y. TIMES (May 6, 2020), <https://www.nytimes.com/2020/05/06/opinion/facebook-oversight-board.html> [<https://perma.cc/PJ52-28RS>].

161. *Oversight Board*, INSTAGRAM, <https://www.instagram.com/oversightboard/> [<https://perma.cc/4RGM-9VCA>].

162. *Oversight Board*, LINKEDIN, <https://www.linkedin.com/company/oversight-board-administration/> [<https://perma.cc/A5VC-VR2E>].

163. @OversightBoard, X (FORMERLY KNOWN AS TWITTER), <https://twitter.com/OversightBoard> [<https://perma.cc/PRR6-Y6FE>].

164. @OversightBoard, X (FORMERLY KNOWN AS TWITTER) (Mar. 2, 2022), <https://twitter.com/OversightBoard/status/1499045053142081543> [<https://perma.cc/6KL7-BJG6>] (“The Oversight Board is closely following the situation in Ukraine and how Meta is responding.”).

165. *See, e.g.*, Cat Zakrzewski, *Facebook Oversight Board Sternly Criticizes the Company’s Collaboration in First Transparency Reports*, WASH. POST (Oct. 21, 2021), <https://www.washingtonpost.com/technology/2021/10/21/facebook-oversight-board-transparency/> [<https://perma.cc/25ZS-7G62>].

166. *See Oversight Board to Meet with Frances Haugen*, OVERSIGHT BD. (Oct. 2021), <https://www.oversightboard.com/news/1232363373906301-oversight-board-to-meet-with-frances-haugen/> [<https://perma.cc/Z3WC-7RTK>].

constantly appear at public events, attending events such as the U.N. General Assembly,¹⁶⁷ the Aspen Ideas Festival,¹⁶⁸ the Paris Peace Forum,¹⁶⁹ and annual Austin festival SXSW.¹⁷⁰ Members give interviews to journalists extolling their work for feature stories.¹⁷¹ When a U.S. district court issued a ruling about government communications with social media platforms, two Board members wrote opinion pieces in mainstream outlets explaining how the Board had dealt with a case raising the same issues (and recommending a better solution).¹⁷² The Board has a newsletter promoting its activities and releases quarterly transparency reports extolling its achievements in getting Meta to enact reforms.¹⁷³

Through these and similar activities, the Board is making an argument about its purpose and utility. In academic work, one of the Board's original Co-Chairs, Professor Jamal Greene, wrote that in traditional legal systems “the practice of constitutional law is the practice of persuading diverse citizens to share the priorities of the adjudicator.”¹⁷⁴ Professor Greene clearly believes persuasion to be an important part of governance, and this is in keeping with the Board's public relations efforts. But the Board is approaching this task in ways that are not typical of what one might expect from an august “Supreme Court”-like body.

167. *Oversight Board*, LINKEDIN (Sept. 24, 2022), https://www.linkedin.com/posts/oversight-board-administration_unga77-oversightboard-unga-activity-6980194741380337666-DNsk [<https://perma.cc/A4JP-W3MG>]; @DigitalRightsPK, X (FORMERLY KNOWN AS TWITTER) (Nov. 25, 2022), <https://twitter.com/DigitalRightsPK/status/1596105015244066816> [<https://perma.cc/A36P-856Q>].

168. The Aspen Institute, *What to Take Down, What to Leave Up, and Why*, YOUTUBE (June 27, 2023), <https://www.youtube.com/watch?v=M1apIrrItg> [<https://perma.cc/2JS2-GTZK>].

169. *The Oversight Board*, PARIS PEACE FORUM (Nov. 2022), <https://parispeaceforum.org/projects/the-oversight-board/> [<https://perma.cc/ZS38-9BE5>].

170. *Lawful, but Awful: Curbing Harmful Content Online*, SXSW 2022 SCHEDULE (Mar. 12, 2022), <https://schedule.sxsw.com/2022/events/PP117894> [<https://perma.cc/J8MS-R9TW>].

171. *E.g.*, Steven Levy, *Inside Meta's Oversight Board: 2 Years of Pushing Limits*, WIRED (Nov. 8, 2022), <https://www.wired.com/story/inside-metas-oversight-board-two-years-of-pushing-limits/> [<https://perma.cc/NT4W-TTSX>].

172. Michael McConnell, *Why Are Government Social Media Takedown Requests Secret? Make Them Public*, WASH. POST (July 7, 2023), <http://www.washingtonpost.com/opinions/2023/07/07/government-social-media-takedown-requests-public/> [<https://perma.cc/N2CV-HEEX>]; Suzanne Nossel, *Banning Government Officials from Talking to Big Tech is No Win for Free Speech*, L.A. TIMES (July 9, 2023), <https://www.latimes.com/opinion/story/2023-07-09/government-social-media-censorship-free-speech> [<https://perma.cc/P89F-BBGM>].

173. *See, e.g.*, *Q2 2023 Transparency Report: Board's Recommendations Lead to Key Changes in Meta's Cross-Check Program*, OVERSIGHT BD. (Oct. 2023), <https://oversightboard.com/news/228158946731169-q2-2023-transparency-report-board-s-recommendations-lead-to-key-changes-in-meta-s-cross-check-program/> [<https://perma.cc/W88F-ZDPJ>]. *News and Articles*, OVERSIGHT BD., <https://www.oversightboard.com/news/> [<https://perma.cc/KY7N-PH5N>].

174. Jamal Greene, *Pathetic Argument in Constitutional Law*, 113 COLUM. L. REV. 1389, 1454 (2013).

Courts do not typically have Twitter accounts that retweet compliments of them, nor do they use the royal “we” to boast that “the results we obtained so far show we are making progress.”¹⁷⁵ Board representatives rarely (if ever) prominently discuss the obvious shortcomings of its institutional design that hamstringing its ability to be a comprehensive form of oversight. Instead, the Board has become a politically active institution, aggressively promoting its own virtues and acting sometimes more like an influencer than a governing body. It insists in its publications, through its members, and through its social media presence that it is performing well and deserves to be at the center of discourse about content moderation governance.

3. Prompting Reforms to Meta’s Systems

The Board has aggressively exercised its power under the Charter to provide optional policy advisory statements in addition to individual decisions in each case,¹⁷⁶ giving multiple such recommendations in every decision. Critics previously suggested that the Board could only have limited impact on Meta because the format of individual appeals meant that the Board would only review a small fraction of Meta’s decisions.¹⁷⁷ The Board’s practice of issuing sweeping recommendations that target Meta’s underlying systems has been its answer to this critique.

It has emphasized the importance of this policy guidance to its work as an institution, insisting its “top priority” is to make sure that Meta implements the Board’s recommendations.¹⁷⁸ The Board issues quarterly transparency reports tracking the comprehensiveness of Meta’s responses to every single recommendation the Board has made.¹⁷⁹ The “non-binding” status of recommendations means that the Board has even less authority to mandate implementation of them than it does to mandate compliance with its individual case decisions. But notwithstanding this fact, the Board itself says that Meta has made progress in implementing the vast majority of reforms the Board has suggested.¹⁸⁰

In reality, the situation is more complex than the Board describes. Although the headline figures are impressive (Meta has only declined

175. @OversightBoard, X (FORMERLY KNOWN AS TWITTER) (Nov. 14, 2022), <https://twitter.com/OversightBoard/status/1587553721826906113> [<https://perma.cc/CRS8-NRCX>].

176. CHARTER, *supra* note 51, art. 1, § 4.

177. See Klonick, *The Facebook Oversight Board*, *supra* note 44, at 2490.

178. 2021 ANNUAL REPORT, *supra* note 112, at 60.

179. See, e.g., *Q1 2023 Transparency Report: Board Publishes New Data on the Impact of Its Recommendations*, OVERSIGHT BD. (June 2023), <https://oversightboard.com/news/1008878700278435-q1-2023-transparency-report-board-publishes-new-data-on-the-impact-of-its-recommendations/> [<https://perma.cc/2JJG-AJ5T>].

180. Q1 2023 QUARTERLY TRANSPARENCY REPORT, *supra* note 133, at 18.

to implement fifty out of 251 recommendations!¹⁸¹), there is a dramatic variance in the importance of the reforms that the Board has been responsible for precipitating. Some of the reforms the Board has pushed Meta to make have been in response to fairly obvious problems. That the Board's push was necessary to make the company adopt them speaks more to Meta's negligence rather than the Board's vigilance. For example, the Board recommended that Meta translate its Community Standards into Punjabi, one of the most-spoken languages in the world.¹⁸² Such a recommendation hardly requires expertise in freedom of expression to recognize as important, but the Board highlighting the issue does appear to have made Meta act more quickly than it would have otherwise. In another case, the Board found that one of Meta's internal policies wasn't being enforced because it was lost for three years.¹⁸³ Again, it is hardly a profound insight that this is not best practice to misplace internal guidance documents. The Board has an obvious and positive impact in cases like these, but perhaps not the kind of impact that should really require years and \$280 million to bring about. In such cases, the Board is playing the role of a glorified auditor rather than an appellate review body. Identifying points of failure can be important but is not necessarily the original vision for the institution.

Other reforms sound good, but their impacts are unclear. For example, a common recommendation in the Board's decisions is that Meta should add further language and clarity to its various Community Standards.¹⁸⁴ The Board asserts that these recommendations are designed to further the IHRL principle of "legality" (that requires rules to be clear and accessible), but whether they actually make a difference to users or make content moderation any more predictable and consistent is an open question.¹⁸⁵ As I will return to below, these kinds of recommendations evince a formalism in the Board's thinking that reflects

181. OVERSIGHT BD., TRANSPARENCY REPORT: SECOND HALF OF 2023 4 (2024), <https://oversightboard.com/attachment/1108810987097200/> [https://perma.cc/YVB7-M2GB].

182. Punjabi Concern Over the RSS in India Case, 2021-003-FB-UA (Oversight Bd. Apr. 29, 2021), <https://www.oversightboard.com/decision/FB-H6OZKDS3/> [https://perma.cc/MJ4V-KQDA]; *Case on Punjabi Concern over the RSS in India*, META TRANSPARENCY CTR. (July 13, 2022), <https://transparency.fb.com/oversight/oversight-board-cases/punjabi-concern-over-the-rss-in-india/> [https://perma.cc/M2QL-YBLM].

183. See Öcalan's Isolation Case, 2021-006-IG-UA (Oversight Bd. July 8, 2021), <https://www.oversightboard.com/decision/IG-I9DP23IB> [https://perma.cc/3BRH-P36B] ("The Board is concerned that Facebook *misplaced* an internal policy exception for three years and that this may have led to many other posts being wrongly removed." (emphasis added)).

184. See, e.g., OVERSIGHT BD., OVERSIGHT BOARD Q1 2022 TRANSPARENCY REPORT 21–22 (2022).

185. See, e.g., Evelyn Douek, *The Siren Call of Content Moderation Formalism*, in SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY 139, 145–49 (Lee Bollinger & Geoffrey Stone eds., 2022) (presenting arguments that these kinds of recommendations are futile).

legalistic intuitions about the nature of rule-making rather than empirically demonstrated impact.¹⁸⁶

But there are certain reforms that the Board has brought about that indisputably represent meaningful demonstrations of authority, too. For example, after long-standing concerns about Meta's enforcement practices in Israel and Palestine and potential disparate impacts on different populations,¹⁸⁷ the Board prompted Meta to engage an independent entity to review its operations in the region, leading to a human rights impact assessment (which concluded that unintentional bias had led to adverse impacts on Palestinian and Arabic speaking users).¹⁸⁸ Meta also committed, in response to Board recommendations, to begin notifying users if their content is removed due to an extra-legal request from a government actor and to include information in its transparency reporting about how many such requests it receives.¹⁸⁹ Both of these initiatives are examples of reforms in areas that civil society had long been raising concerns about for years without gaining traction.¹⁹⁰ This suggests that the Board has some meaningful capacity to influence Meta that other outsiders lack.

The Board's recommendations have also become more sophisticated and targeted over time. One of the biggest areas of improvement for the Board over the course of its first term has been the shift in its thinking from a highly individualistic analysis of the issues before it to more systemic thinking. That is to say, the Board has moved from very case-specific, fact-intensive decision-making,¹⁹¹ to decisions that are more focused on Meta's underlying systems for content moderation

186. *Infra* Section III.B.2.

187. See, e.g., JILLIAN C. YORK, SILICON VALUES: THE FUTURE OF FREE SPEECH UNDER SURVEILLANCE CAPITALISM 58 (2021).

188. Dunstan Allison-Hope, Jenny Vaughan & Lindsey Andersen, *Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021*, BSR (Sept. 22, 2022), <https://www.bsr.org/en/blog/human-rights-due-diligence-of-meta-impacts-in-israel-and-palestine-may-2021> [<https://perma.cc/5A46-6UMN>].

189. *Case Regarding the Support of Abdullah Öcalan, Founder of the PKK*, META TRANSPARENCY CTR. (July 13, 2022), <https://transparency.fb.com/oversight/oversight-board-cases/support-of-abdullah-ocalan-founder-of-the-pkk> [<https://perma.cc/276K-A2JK>] (adopting Oversight Board Recommendation 9).

190. See, e.g., *Open Letter to Facebook, Twitter, and YouTube: Stop Silencing Critical Voices from the Middle East and North Africa*, ACCESS NOW (Dec. 17, 2020), <https://www.accessnow.org/press-release/facebook-twitter-youtube-stop-silencing-critical-voices-mena/> [<https://perma.cc/SF2U-PXTM>] ("Palestinian activists and social media users have been campaigning since 2016 to raise awareness around social media companies' censorial practices."); Scott Craig & Emma Llansó, *Pressuring Platforms to Censor Content is Wrong Approach to Combatting Terrorism*, CTR. FOR DEMOCRACY & TECH. (Nov. 5, 2015), <https://cdt.org/insights/pressuring-platforms-to-censor-content-is-wrong-approach-to-combatting-terrorism/> [<https://perma.cc/LZK7-PF9>] ("Company transparency reporting can also help to provide more information about content removal requests from governments . . .").

191. See Evelyn Douek, *The Facebook Oversight Board's First Decisions: Ambitious, and Perhaps Impractical*, LAWFARE (Jan. 28, 2021), <https://www.lawfareblog.com/facebook-oversight-boards-first-decisions-ambitious-and-perhaps-impractical> [<https://perma.cc/VS8S-X648>] (critiquing the Board's early decisions along these lines).

enforcement and platform design choices.¹⁹² Early decisions turned on matters like comparing slightly different translations of individual posts,¹⁹³ or different inferences about the poster’s individual intent.¹⁹⁴ Acknowledging the impossibility of getting every decision right at scale, the Board’s later decisions move away from such fact-specific analysis, and instead they look for discriminatory patterns of enforcement¹⁹⁵ and use individual errors to identify “systemic breakdown[s]” requiring reform.¹⁹⁶ Because of the trade-offs and complexities created by the volume and speed of content moderation, this move from an individualistic to a more systems-thinking-based approach is necessary to address the system design choices that matter at scale.¹⁹⁷ The shift in the Board’s thinking is marked, showing growth in members’ understanding of the systems they are overseeing and increasing their potential for impact.

There have also been some significant losses for the Board, however, including (or perhaps especially) in high-profile cases. Meta refused to take up in any meaningful way one of the most consequential recommendations in the *Trump Suspension Case*, to “review its potential role in the election fraud narrative that sparked violence in the United States on January 6, 2021 and report on its findings.”¹⁹⁸ It similarly did not adopt the Board’s recommendation to provide a specific transparency report about Community Standards enforcement during the COVID-19 pandemic.¹⁹⁹ And it withdrew a request for policy advice concerning content moderation issues related to Russia’s ongoing

192. *See, e.g.*, Wampus Belt Case, 2021-012-FB-UA (Oversight Bd. Dec. 9, 2021), <https://www.oversightboard.com/decision/FB-L1LANIA7/> [<https://perma.cc/X9VU-T5NY>] (discussing the importance of systems and design); Colombian Police Cartoon Case, 2022-004-FB-UA (Oversight Bd. Sept. 15, 2022), <https://www.oversightboard.com/decision/FB-I964KKM6> [<https://perma.cc/VC7C-FCPP>].

193. *See, e.g.*, Myanmar Post About Muslims Case, 2020-002-FB-UA (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/FB-I2T6526K> [<https://perma.cc/3RNP-4E38>].

194. *See, e.g.*, “Two Buttons” Meme Case, 2021-005-FB-UA (Oversight Bd. May 20, 2021), <https://www.oversightboard.com/decision/FB-RZL57QHJ> [<https://perma.cc/JV77-X4WN>].

195. *See, e.g.*, Reclaiming Arabic Words Case, 2022-003-IG-UA (Oversight Bd. June 13, 2022), <https://www.oversightboard.com/decision/IG-2PJ00L4T> [<https://perma.cc/CW6L-DFV2>].

196. *See, e.g.*, Knin Cartoon Case, 2022-001-FB-UA (Oversight Bd. June 17, 2021), <https://oversightboard.com/decision/FB-JRQ1XP2M/> [<https://perma.cc/3A3W-B86D>].

197. Douek, *supra* note 127, at 548–56.

198. *Case on Former President Trump’s Suspension from Facebook*, META TRANSPARENCY CTR. (June 12, 2023), <https://transparency.fb.com/oversight/oversight-board-cases/former-president-trump-suspension-from-facebook> [<https://perma.cc/AU6P-S9SM>].

199. Evelyn Douek, *The Oversight Board Moment You Should’ve Been Waiting For: Facebook Responds to the First Set of Decisions*, LAWFARE (Feb. 26, 2021), <https://www.lawfareblog.com/oversight-board-moment-you-shouldve-been-waiting-facebook-responds-first-set-decisions> [<https://perma.cc/JZ8P-GME5>].

war with Ukraine, against the Board’s wishes.²⁰⁰ That is, in three of the most controversial and consequential content moderation subject areas in the last few years, the limits of the Board’s power over Meta were on stark display. In other less visible cases, Meta “repeatedly push[es] back the deadline for implementation,” and the Board is unable to do anything other than highlight Meta’s slow walking.²⁰¹

And once again, it is important to be aware of the underlying incentives at play when interpreting the data that the Board and Meta release. Both organizations have good reason to paint as glowing a picture of the Board’s accomplishments as they can, in order to try to convince outsiders of the benefits of the Board as an institution and reap the legitimacy dividends. This is why it is important to remain skeptical about Meta’s claims about the Board’s impacts. Some of the changes Meta says it has made in response to the Board’s recommendations may have been made regardless. Moreover, it is not always obvious what it means when the company says it is implementing the Board’s recommendations in part or in full. Originally, Meta also counted itself as “committed to action” in response to the Board’s recommendations when its only response was to point to actions it was already taking.²⁰² After I criticized this as inflating the impact of the Board,²⁰³ Meta implicitly acknowledged this was misleading by introducing a new category of response, “[w]ork Meta already does.”²⁰⁴ But Meta’s categorization of its responses often remains very generous. For example, when the Board recommended that Meta “increase its investments in digital literacy programs,” Meta classified itself as implementing the Board’s recommendation “in [f]ull” because it was committing to “continue to globally roll-out” its existing programs²⁰⁵ — a commitment that seems unlikely to have been the “increase” in investment the Board had in mind.

200. *Meta Withdraws a Policy Advisory Opinion Request Related to Russia’s Invasion of Ukraine*, META TRANSPARENCY CTR. (June 12, 2023), <https://transparency.fb.com/oversight/oversight-board-cases/ukraine-russia-pao> [<https://perma.cc/S4M9-S87E>]; *Protecting Freedom of Expression and Human Rights in Ukraine and Russia*, OVERSIGHT BD. (May 2022), <https://www.oversightboard.com/news/382264103827624-protecting-freedom-of-expression-and-human-rights-in-ukraine-and-russia/> [<https://perma.cc/7YXS-2K2Z>].

201. 2022 ANNUAL REPORT, *supra* note 21, at 24.

202. *Case on Breast Cancer Symptoms and Nudity*, OVERSIGHT BD. (June 12, 2023), <https://transparency.meta.com/oversight/oversight-board-cases/breast-cancer-symptoms-nudity/> [<https://perma.cc/TH7N-YJZD>]; *see also* *Breast Cancer Symptoms and Nudity Case*, 2020-004-IG-UA (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/IG-7THR3SII> [<https://perma.cc/LEU7-H9HH>] (the underlying case decision).

203. Douek, *supra* note 199.

204. *Oversight Board Recommendations*, META TRANSPARENCY CTR., <https://transparency.fb.com/pt-br/oversight/oversight-board-recommendations/> [<https://perma.cc/4S3B-JN4A>].

205. *Oversight Board Selects a PAO on the Removal of COVID-19 Misinformation*, META TRANSPARENCY CTR. (June 20, 2023), <https://transparency.fb.com/oversight/oversight-board-cases/pao-on-covid-19-misinformation> [<https://perma.cc/4E4X-TSFU>].

As I discuss below,²⁰⁶ assessing the tangible impact of the reforms that the Board has brought about is a much trickier — and much more important — question than the mere fact that the Board has prompted Meta to initiate reforms. But the basic fact remains that the Board, an institution with no formal legal power, has succeeded in making one of the most powerful corporations in the world change the way it operates.

4. Making Meta More Transparent and Accountable

The Board has brought some (albeit limited) forms of transparency and accountability to Meta's content moderation that, despite being perhaps the most pervasive system of speech regulation in history, was previously almost entirely opaque. When the Board selects a case, it sends Meta a series of questions, often about the relevant enforcement processes and internal rules, and Meta has answered the vast majority.²⁰⁷ The Board sometimes includes parts of Meta's answers in its decisions, and through this process the Board has revealed new information about the company in a range of areas: from details about Meta's database of images that automatically get removed from its platforms,²⁰⁸ to the quality assurance systems the platform has in place,²⁰⁹ to the way the company makes allowances for newsworthy content in moments of political protest.²¹⁰ Thus, the Board has imposed accountability at least in the thin sense of requiring Meta to explain itself.²¹¹ The Board has also imposed accountability through third parties by, for example, prompting Meta to undertake a human rights impact assessment in Palestine.²¹² In these ways, the Board has helped shed light on systems that were previously completely dark to outsiders and forced Meta to explain (and often rethink) its decisions.

But there have also been real limits to how much accountability the Board has been able to impose, not least the fact that the Board still has jurisdiction over only the tiniest fraction of decisions that Meta

206. See *infra* Section III.B.2.

207. 2022 ANNUAL REPORT, *supra* note 21, at 38 (reporting that Meta answered eighty-six percent of the Board's questions fully and that the answers were increasingly comprehensive).

208. Colombian Police Cartoon Case, 2022-004-FB-UA (Oversight Bd. Sept. 15, 2022), <https://www.oversightboard.com/decision/FB-I964KKM6> [<https://perma.cc/VC7C-FCPP>].

209. See, e.g., Mention of the Taliban in News Reporting Case, 2022-005-FB-UA (Oversight Bd. Sept. 15, 2022), <https://www.oversightboard.com/decision/FB-U2HHA647> [<https://perma.cc/2H5T-L5WL>]; Meta's Cross-Check Program Policy Advisory Opinion, PAO-2021-02 (Oversight Bd. Dec. 6, 2022), <https://www.oversightboard.com/decision/PAO-NR730OF1/> [<https://perma.cc/K7Z3-PCK6>].

210. Iran Protest Slogan Case, 2022-013-FB-UA (Oversight Bd. Jan. 9, 2023), <https://www.oversightboard.com/decision/FB-ZT6AJS4X> [<https://perma.cc/S2S9-YUHN>].

211. This definition of accountability draws on Bamberger, *supra* note 95, at 404; Freeman, *supra* note 95, at 664; Mashaw, *supra* note 95, at 117.

212. See Allison-Hope et al., *supra* note 188.

makes.²¹³ Meta can also refuse to answer the Board’s questions for any of the broad reasons given in the Bylaws (such as if Meta “determines that the information is not reasonably required”)²¹⁴, and the Board cannot contest Meta’s determinations or compel it to disclose anything it doesn’t want to.²¹⁵ In some cases, it has later become clear that Meta was withholding relevant information or selectively disclosing information to the Board in ways that were misleading.²¹⁶ In some cases, Meta has also rejected Board recommendations that would have led to greater accountability, such as investigations into its content moderation during the pandemic or in the lead up to January 6,²¹⁷ and has simply withdrawn cases from the Board’s review on sensitive topics, with little explanation.²¹⁸

Even in these cases though, the Board has provided a focal point for other stakeholders’ demands for accountability. For example, lawmakers have pointed to Meta’s refusal to implement recommendations from “its own Oversight Board” as a sign of the company’s failures.²¹⁹ They have also used the Board’s decisions as moments of leverage to encourage Meta to enact reforms.²²⁰ And the fact that Meta had misled the Board on its Cross-Check program for high-profile users was

213. See Douek, *supra* note 57. The Board’s jurisdiction has been marginally expanded since that piece was written, to include cases where individuals are appealing for content to be taken down, but the overall point still stands. Rosen, *supra* note 85.

214. BYLAWS, *supra* note 51, art. 2, § 2.2.2.

215. See Arun, *supra* note 109, at 261–62.

216. See Catalina Botero-Marino, Jamal Greene, Michael McConnell & Helle Thorning-Schmidt, *To Treat Users Fairly, Facebook Must Commit to Transparency*, OVERSIGHT BD. (Sept. 2021), <https://www.oversightboard.com/news/3056753157930994-to-treat-users-fairly-facebook-must-commit-to-transparency/> [<https://perma.cc/8LBA-SDC8>] (noting that Meta had not been fully forthcoming about its Cross-Check program); Evelyn Douek, *Facebook’s Responses in the Trump Case Are Better Than a Kick in the Teeth, but Not Much*, LAWFARE (June 4, 2021), <https://www.lawfareblog.com/facebooks-responses-trump-case-are-better-kick-teeth-not-much> [<https://perma.cc/4K5P-YQNH>] (noting that Meta gave the Board incorrect answers about how many times it had applied its newsworthiness policy); Sam Schechner, *Facebook Is Rebuked by Oversight Board Over Transparency on Treatment of Prominent Users*, WALL ST. J. (Oct. 21, 2021), <https://www.wsj.com/articles/facebooks-oversight-board-says-company-wasnt-fully-forthcoming-on-treatment-of-high-profile-users-11634817601> [<https://perma.cc/P7QP-X5CG>].

217. Douek, *supra* note 199.

218. See, e.g., *Meta Withdraws a Policy Advisory Opinion Request*, *supra* note 200.

219. See Press Release, Sen. Amy Klobuchar, Approaching the Anniversary of January 6th Insurrection, Klobuchar, Colleagues Press Meta About Failure to Address Election-Related Misinformation and Disinformation (Dec. 22, 2021), <https://www.klobuchar.senate.gov/public/index.cfm/2021/12/approaching-the-anniversary-of-january-6th-insurrection-klobuchar-colleagues-press-meta-about-failure-to-address-election-related-misinformation-and-disinformation> [<https://perma.cc/F6MK-YKXC>].

220. Press Release, Rep. Lori Trahan, Trahan, Schiff Respond to Meta Oversight Board’s Recommendation to Overhaul Cross-Check Program (Dec. 6, 2022), <https://trahan.house.gov/news/documentsingle.aspx?DocumentID=2690> [<https://perma.cc/8FBA-QVAY>].

prominently reported on,²²¹ which was no doubt a significant reason Meta later referred the Cross-Check program to the Board.²²²

Thus, the scope of accountability that the Board can impose is limited (only a fraction of the kinds of decisions Meta makes), the kind of accountability thin (simply the requirement for Meta to explain itself), and the consequences for transgression slight (a public rebuke). That said, the Board has proven that even this form of accountability has value. The Board has forced Meta to provide information and justification for decisions that previously went entirely unexamined and helped create further levers for stakeholders to pressure the company. This has positive externalities for broader debates about content moderation as the complexities and trade-offs inherent in running a system of speech regulation at the speed and size of a major platform are opened up for public examination. To the extent that the Board's review process often reveals previously undiscovered or unacknowledged mistakes on Meta's behalf, which it then corrects,²²³ this also shows a form of accountability and the benefits of Meta being forced to explain itself, even to a minimal degree.²²⁴

5. Making Content Moderation More Democratic

For some, the Board represented an opportunity to make “digital spaces more democratic” and “one of the first attempts to open up the decision-making system of a commercial platform to the ‘outside.’”²²⁵ But what does it mean to make content moderation “more democratic”? While a decision will generally be considered “democratically responsive to the extent it reflects and expresses the popular will,”²²⁶ the way that the popular will should be distilled and translated into specific

221. See, e.g., Schechner, *supra* note 216 (describing the Cross-Check program as one “initially intended as a quality-control measure for actions taken against high-profile accounts [but that] has grown to include millions of accounts.”).

222. Nick Clegg, *Requesting Oversight Board Guidance on Our Cross-Check System*, META (Sept. 28, 2021), <https://about.fb.com/news/2021/09/requesting-oversight-board-guidance-cross-check-system/> [<https://perma.cc/N3TH-Z5HC>].

223. See, e.g., Q1 2023 QUARTERLY TRANSPARENCY REPORT, *supra* note 133, at 9 (“[I]t is noted that Meta found its original decision to have been incorrect in 63% of cases the Board shortlisted The Board continues to raise with Meta the questions this poses for the accuracy of the company’s content moderation and the appeals process the company applies before cases reach the Board.”).

224. For further exploration of this point, see Glen Staszewski, *Reason-Giving and Accountability*, 93 MINN. L. REV. 1253, 1280 (2009) (“Because public officials must provide public-regarding justifications for their decisions, other participants in the process have incentives to articulate their claims in public-regarding terms as well. As a result, relatively selfish policy options may be discarded . . .”).

225. Kettemann & Schulz, *supra* note 109, at 6.

226. Nina A. Mendelson, *Rulemaking, Democracy, and Torrents of E-Mail*, 79 GEO. WASH. L. REV. 1343, 1350 (2011).

content moderation decisions is usually left under-specified.²²⁷ It is impossible to say whether the Board has made content moderation more democratic without a more particularized conception of what that would mean. But it is at least true that it has provided an avenue for a broader set of stakeholders to comment on Meta’s rules and decisions.

The Board has broadly interpreted its discretion to “request and receive information from a global pool of outside subject-matter experts”²²⁸ to allow it to issue a call for public comment at the time it announces that it has taken a new case. The Board has called this process “crucial” for achieving its goals of improving Meta’s processes and stated that “[o]n numerous occasions, public comments have shaped our decisions and our recommendations to Meta.”²²⁹ Through this mechanism, then, the Board has created an avenue for anyone and everyone to express their views on Meta’s decisions. To the extent that calls for democratization simply reflect the desire for broader participation in content moderation policy and decisions, the Board could, for this reason, be considered successful. Indeed, as noted earlier, a wide set of stakeholders — including leading civil society organizations, academics, and politicians — have taken advantage of this process.²³⁰ And it is not just elites — the broader public has sometimes also engaged with the Board in this way. In the *Trump Suspension Case*, people seized the opportunity — the Board received 9,666 public comments — but this remains a significant outlier in terms of volume of engagement.²³¹

While the Board is therefore creating a venue for more people than ever to have their views on specific content moderation issues heard, the creation of a public comment process is also not inherently or necessarily a pro-democratic force. The long-standing concerns about skewed participation in public comment processes that favor well-resourced or powerful interest groups in the context of the administrative

227. For an important discussion of this point, see Brenda Dvoskin, *Representation Without Elections: Civil Society Participation as a Remedy for the Democratic Deficits of Online Speech Governance*, 67 VILL. L. REV. 447, 448 (2023) (arguing that “prevailing assumptions about civil society participation in content moderation are wrong”); see also Rachel Griffin, *Public and Private Power in Social Media Governance: Multistakeholderism, the Rule of Law and Democratic Accountability*, 14 TRANSNAT’L LEGAL THEORY 46, 69–73 (2023) (discussing the pitfalls of private, market-based governance structures).

228. BYLAWS, *supra* note 51, art. 1, § 3.1.4.

229. 2021 ANNUAL REPORT, *supra* note 112, at 51.

230. See *supra* Part I.

231. See, e.g., *Trump Suspension Case*, *supra* note 1. By way of comparison, in the latter half of 2023, the Board received just 245 public comments *across all cases*. *H2 2023 Transparency Report: How the Board’s Ongoing Impact Is Making a Difference*, OVERSIGHT BD. (Mar. 19 2024), <https://oversightboard.com/news/7451697308245799-h2-2023-transparency-report-how-the-board-s-ongoing-impact-is-making-a-difference/> [<https://perma.cc/3CPN-CV63>].

state²³² apply equally to the Board’s process,²³³ in which wealthier and English-speaking regions comprise the majority of comments.²³⁴

There is also the question of what the Board does with these comments once it receives them.²³⁵ Indeed, whether public comments have any effect (let alone a democratic one) on the Board’s work is hard to tell, especially from the outside. In some cases, public comments appear to have had an impact. In the *Gender Identity and Nudity Cases*,²³⁶ for example, the Board made special note of the high level of public comment, especially from the minority community likely to be particularly affected by the decision (in that case, trans and non-binary people).²³⁷ It was the first time the Board described the way public comments had affected its deliberation in detail, noting that it takes “comments seriously as a part of its deliberations,” while understanding that “these comments may not be representative of global opinion.”²³⁸ In that case, however, the Board found the public comments’ evidence of the disparate impact of Meta’s policies persuasive, and this clearly informed the Board’s decision to recommend that Meta reform its rules.

References to particular public comments seem to be increasing in the Board’s more recent cases, but overall the *Gender Identity and Nudity Cases* are an outlier, a rare instance in which public comment played a significant role in the Board’s explicit reasoning — the vast majority of the time, the Board does not engage with public comments beyond summarizing them at a very high level.²³⁹ Even in the *Trump Suspension Case*, where the Board received nearly ten thousand comments, it is not clear whether these comments affected the Board’s decision in any way.²⁴⁰ The Board did not refer to any such impact, beyond noting that it was “grateful for the many thoughtful and

232. Mendelson, *supra* note 226, at 1357–58 (summarizing the literature on these concerns).

233. See Dvoskin, *supra* note 93, at 9 (“[O]rganizations might not have enough resources to engage in as many participatory mechanisms as they wish. Companies might reach out too often; there might be too many open calls and opportunities for engagement. Civil society might not be able to make use of all these spaces.”).

234. See, e.g., OVERSIGHT BD., OVERSIGHT BOARD Q4 2022 QUARTERLY TRANSPARENCY REPORT 14 (2022).

235. Cf. Mendelson, *supra* note 226, at 1359 (making the same point in the context of administrative agencies).

236. 2022-009-IG-UA and 2022-010-IG-UA (Oversight Bd. June 17, 2023), <https://www.oversightboard.com/decision/BUN-IH313ZHJ/> [<https://perma.cc/ZEP2-87KA>].

237. *Id.*

238. *Id.*

239. E.g., Ayahuasca Brew Case, 2021-013-IG-UA (Oversight Bd. Dec. 9 2021), <https://www.oversightboard.com/decision/IG-0U6FLA5B> [<https://perma.cc/3EZZ-7FBA>] (summarizing public comments in a single paragraph). As Brenda Dvoskin has summarized it, “[d]espite the Board’s interest in stakeholder engagement initiatives, the view from nowhere has, so far, prevailed in the Board’s deployment of the IHRL project.” Brenda Dvoskin, *Expert Governance of Online Speech*, 64 HARV. INT’L L.J. 85, 128 (2023).

240. See *Trump Suspension Case*, *supra* note 1.

engaged public comments that it received.”²⁴¹ (It also did not mention that it probably received many less thoughtful comments, too.)

Nor is it clear how much weight the Board gives public comments compared with the other forms of input the Board solicits. Every decision has a note at the end that the Board commissioned independent research from, most often, the University of Gothenburg. The Board has never revealed the substance of this research or referenced it in a decision. In some cases, it organizes “virtual roundtable[s]” with advocacy groups, presumably hand-picked.²⁴² It has also cited to a single academic article, once.²⁴³ But presumably (hopefully) the Board’s work is also informed by other academic research, even if not referenced in its decisions.

In sum, while it is clear that the Board has created new ways for the public to express opinions about Meta’s rules, how this has altered the power dynamics of decision-making — and whether it has done so for the better — is much more ambiguous. This is not, of course, a problem that is unique to the Board. Many governance regimes face the same challenge, and there is no single model of how best to incorporate public input.²⁴⁴ But the Board could and should be much more transparent about how it makes use of public comments,²⁴⁵ as well as other sources of input, lest ultimately it undermines faith in its own public responsiveness. The Board’s public comment process should not be “to public participation as Japanese Kabuki theater is to human passions — a highly stylized process for displaying in a formal way the essence of something which in real life takes place in other venues.”²⁴⁶ The Board has acknowledged the limits of its own democratic credentials as a hand-picked group of experts, by opening up its cases for public comment, but it should be more consistent about showing that this consultation is genuine and not merely performative.

241. *Id.*

242. *See, e.g.*, Swedish Journalist Reporting Sexual Violence Against Minors Case, 2021-016-FB-FBR (Oversight Bd. Feb. 1, 2022), <https://www.oversightboard.com/decision/FB-P9PR9RSA> [<https://perma.cc/B6E2-627B>].

243. Pro-Navalny Protests in Russia Case, 2021-004-FB-UA (Oversight Bd. May 26, 2021) (citing Professor Kate Klonick), <https://www.oversightboard.com/decision/FB-6YHRXHZR> [<https://perma.cc/WH9R-DNLE>].

244. *See* Katharine Jackson, *What Makes an Administrative Agency “Democratic”?*, LPE PROJECT (Nov. 11, 2020), <https://lpeproject.org/blog/what-makes-an-administrative-agency-democratic/> [<https://perma.cc/MPS6-T36W>].

245. *See* Mendelson, *supra* note 226, at 1379 (“Some accountability could be provided if the agency is required (as it might be in judicial review) to be transparent about its response to significant levels of public comment. The agency could be required to give the comments some consideration and provide its reasons for giving the weight that it did.”).

246. E. Donald Elliott, *Re-Inventing Rulemaking*, 41 DUKE L.J. 1490, 1492 (1992).

6. Legitimacy Building

Perhaps the most important achievement of the Board is the way it has managed to transform the tenor of public engagement with it from highly critical to often accepting, even respectful. I have already noted the media reporting, the extensive stakeholder engagement with its public comment process, and the inclusion of the Board's case decisions in sets normally reserved for formal legal authorities such as legal databases and case notes.²⁴⁷

The academic literature on the Board also takes the institution remarkably seriously. There are numerous articles analyzing its decisions and use of IHRL.²⁴⁸ Scholars (and Board members) have suggested that, through “norm diffusion,” the Board could directly influence the jurisprudence of human rights bodies,²⁴⁹ as well as for national courts and regulators.²⁵⁰ It has been called a source of “[t]ransnational [c]onstitutional [a]dvice”²⁵¹ with a “quality of adjudication” that is “remarkable” and a body which could signal a “new wave of transnational hybrid adjudication.”²⁵² Newton and Martha Minow have pointed to the Board as a salutary guide for effective self-regulation from which the industry more broadly should learn.²⁵³ David Fontana and David Schleicher have suggested that the Board represents an emerging form of governance that “could be commercially valuable” for many other

247. See *supra* Part I.

248. See, e.g., BARATA, *supra* note 157; *Oversight Board Finds a Facebook Rule's Application Violates International Human Rights Law*, *supra* note 29; Pollicino & De Gregorio, *supra* note 94; Wong & Floridi, *supra* note 94.

249. Laurence R. Helfer & Molly K. Land, *The Meta Oversight Board's Human Rights Future*, 44 *CARDOZO L. REV.* 2233, 2285 (2023). For a discussion of the jurisprudential interaction between the Board and IHRL, see Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 *DUKE L. & TECH. REV.* 26, 56–67 (2018); Sejal Parmar, *Facebook's Oversight Board: A Meaningful Turn Toward International Human Rights Standards?*, *JUST SEC.* (May 20, 2020), <https://www.justsecurity.org/70234/facebooks-oversight-board-a-meaningful-turn-towards-international-human-rights-standards/> [<https://perma.cc/J3XW-GYSY>].

250. Catalina Botero Marino & Erik Tuchtfield, *Quasi-Judicial Oversight Mechanisms for Social Platforms – A Conversation with Catalina Botero Marino, Co-Chair of the Oversight Board*, 3 *RECHT & ZUGANG* 254, 257 (2021) (interviewing an Oversight Board Co-Chair); Matija Miloš & Toni Pelić, *Constitutional Reasoning There and Back Again: The Facebook Oversight Board as a Source of Transnational Constitutional Advice*, in *EUROPEAN YEARBOOK OF CONSTITUTIONAL LAW 2021*, at 197, 216 (Jurgen de Poorter, Gerhard van der Schyff, Maarten Stremmer & Maartje De Visser eds., T.M.C. Asser Press 2022) (“National courts may in some cases find these pieces of [Oversight Board] interpretation persuasive, although not precedential.”).

251. Miloš & Pelić, *supra* note 250, at 197.

252. Rishi Gulati, *Meta's Oversight Board and Transnational Hybrid Adjudication – What Consequences for International Law?*, 24 *GERMAN L.J.* 473, 490 (2023); see also Angelo Jr Golia, *The Transformative Potential of Meta's Oversight Board: Strategic Litigation Within the Digital Constitution?*, 30 *IND. J. GLOB. LEGAL STUD.* 325 (2023).

253. Newton Minow & Martha Minow, Response, *Social Media Companies Should Pursue Serious Self-Supervision – Soon: Response to Professors Douek and Kadri*, 136 *HARV. L. REV. F.* 428, 441–42 (2023).

private firms, far beyond the social media industry, to replicate.²⁵⁴ Laurence Helfer and Molly Land have compared the Board to an international human rights tribunal and argued that the Board can, and already has, “serve[d] as an important check on Meta and [] significantly advance[d] the promotion and protection of rights online.”²⁵⁵ Jonathan Zittrain has cited the Board as an experiment of the kind necessary to bring greater process and legitimacy to digital governance.²⁵⁶ Chinmayi Arun has noted the Board’s limitations but concludes that the Board is meaningfully independent, “highly influential” and plays a “very significant role” in bringing accountability to Meta.²⁵⁷ The Board gets lots of attention from the blogosphere,²⁵⁸ and still one influential commentator has declared that the Board “has not gotten nearly the credit it deserves.”²⁵⁹ I myself have added and am right now adding to the pile of articles about the body.²⁶⁰ The literature is, as they say, burgeoning.

Naturally, not all reviews of the Board’s work are positive. But all influential institutions are of course subject to critique, and the tenor of even many of the critical assessments of the Board have changed from being downright dismissive to often being serious and sincere.²⁶¹

Legitimacy is hard to define and even harder to measure. Still, this kind of serious engagement from a wide diversity of stakeholders surely suggests that the institution is regarded with some respect — at least enough to make the cost of engagement worthwhile.²⁶² For a fledgling institution, created by a corporation that was itself at the peak

254. Fontana & Schleicher, *supra* note 114, at 10.

255. Helfer & Land, *supra* note 130, at 2298.

256. Jonathan Zittrain, *Three Eras of Digital Governance* (2019), <https://www.ssrn.com/abstract=3458435> [<https://perma.cc/B5PD-M22W>].

257. Arun, *supra* note 109, at 262–63.

258. See, e.g., Alexa Koenig, *Meta’s Oversight Board Recommends Major Advance in International Accountability*, JUST SEC. (June 22, 2023), <https://www.justsecurity.org/87015/metasp-oversight-board-recommends-major-advance-in-international-accountability/> [<https://perma.cc/FF5Z-5H7W>]; Jillian C. York & Dia Kayyali, *The Facebook Oversight Board is Making Good Decisions- but Does It Matter?*, TECH POL’Y PRESS (July 28, 2021), <https://techpolicy.press/the-facebook-oversight-board-is-making-good-decisions-but-does-it-matter/> [<https://perma.cc/QLL3-ZDEM>]; Vicki Jackson & Martha Minow, *Facebook Suspended Trump. The Oversight Board Shouldn’t Let Him Back.*, LAWFARE (Mar. 8, 2021), <https://www.lawfaremedia.org/article/facebook-suspended-trump-oversight-board-shouldnt-let-him-back> [<https://perma.cc/ASK6-7V96>]; Jacob Schulz, *Twitter is the (Very Short Term) Winner of the Trump Oversight Board Saga*, LAWFARE (June 10, 2021), <https://www.lawfaremedia.org/article/twitter-very-short-term-winner-trump-oversight-board-saga> [<https://perma.cc/EN62-6LLE>]; Parmar, *supra* note 249.

259. Ben Thompson, *Apple and Property Rights, on Google’s Non-Response, the Facebook Oversight Board*, STRATECHERY (Aug. 30, 2022), <https://stratichery.com/2022/apple-and-property-rights-on-googles-non-response-the-facebook-oversight-board/> [<https://perma.cc/PM59-5KDL>].

260. I’ll let you be the judge of whether that bolsters its legitimacy or not.

261. There remain exceptions. See, e.g., Ari Ezra Waldman, *Disorderly Content*, 97 WASH. L. REV. 907, 969–72 (2022).

262. Price & Price, *supra* note 109, at 3322.

of a legitimacy crisis of its own, this is a somewhat remarkable achievement.

B. The Board's Formalistic Approach

The Board's establishment of legitimacy is made more remarkable by the fact that it has been neglecting some core aspects of its role. While the sections above noted many limits on the Board's success stories due to its institutional design, this Section explores the ways that the Board has under-delivered, even taking its institutional constraints as a given. While the structural critiques of the Board are important, it is vital to appreciate the Board's missed opportunities within its current design because ultimately, as Part IV below will argue, these failures are the result of institutional incentives and political dynamics that apply much more broadly than to the Board alone. That is, if the nature and source of these failures are not adequately appreciated, they are likely to be replicated by other content moderation governance institutions, even if they do not have the same structural weaknesses.

This Section focuses on three key ways in which the Board has failed to fulfill its potential. First, the Board has neglected the task of developing a coherent normative framework for its work. Its decisions mimic the forms of judicial opinion-writing, but the Board's reason-giving has been relatively superficial and failed to engage with the most difficult questions that content moderation governance raises. Second, the Board has not clearly defined its goals, and the metrics that it measures as indicators of its impact are poor proxies for material benefits. It is thus impossible to tell the real-world impacts of the Board's work, and therefore meaningful accountability — for either Meta or the Board — remains illusory. Finally, in the most difficult cases, the Board often avoids giving a clear answer, refusing to take a clear stance on some of the most controversial questions that content moderation raises. The result is an institution that observes many of the formalities of good governance but pays less attention to its substantive outcomes.

1. Rulings Without Reason

The Board's function as a reason-giving institution is core to its mission — its job is not only to decide cases but to publicly explain its decisions.²⁶³ This reflects the aspiration that “[t]he principle of publicity — or public reason-giving — allows for notice, guidance, and prediction, all essential to the rule of law.”²⁶⁴ But despite the centrality of

263. See *supra* Section II.B; see also CHARTER, *supra* note 51, at 2 (“The board will operate transparently and its reasoning will be explained clearly to the public . . .”).

264. Erin F. Delaney, *Analyzing Avoidance: Judicial Strategy in Comparative Perspective*, 66 DUKE L.J. 1, 13 (2016).

reason-giving to the Board's design and purpose, the Board's reasoning has been thin. It has prioritized form over substance, simplicity over depth, and failed to seriously grapple with the most significant questions it was set up to answer.

The Board's reasons have mechanistically followed a uniform template. They begin with a recitation of the relevant facts and submissions, followed by citations to the general rules the Board will apply. Then the opinions evaluate the merits of Meta's original decision under Meta's policies, values, and what the Board considers its human rights responsibilities to be. Finally, there are the Board's conclusions and recommendations. The format never changes, and the style is dry and rote, with the Board preferring clarity and consistency over rhetoric. The decisions are also economical — the analysis has grown gradually longer but remains brief by the standards of most legal decisions, especially ones of such complexity and consequence. There are no dissents. To the extent that any Board members disagree with a decision, this is usually briefly noted at a high level in a sentence or two. As a matter of aesthetic impression, then, the Board's decisions exude competence and professionalism perhaps, but not panache.

The formal veneer of these decisions obscures the ways in which the Board has neglected deeper questions about its role and the basis of its decision-making. Maybe the hope is that if the decisions look like rigorous legal decisions, that's enough. The Board has opted for simplicity and consensus, rather than complexity and argumentation, leaving it with relatively thin theoretical foundations for its work. This makes the Board's reasoning ultimately unsatisfying. The Board performs the role of reason-giver but often gives very little by way of meaningful justification for its decisions.

The most significant manifestation of this is in the Board's development (or lack thereof) of its analytical framework. A central problem for content moderation decision-makers in search of legitimacy is that there is no existing normative framework for their decisions. No prior body of precedent deals with how corporate speech regulators operating at massive scale should govern platforms that are privately owned but significantly affect the public interest.²⁶⁵ Indeed, the Board was created in part because there were no established norms, agreed modes of analysis, or preexisting public forums for developing a coherent body of well-reasoned and principled content moderation decisions. The Board's public reason-giving was therefore intended to fill this gap by developing a normative framework upon which to base private platform speech regulation.

265. See, e.g., *Packingham v. North Carolina*, 582 U.S. 98, 107 (2017) (observing that social media platforms “for many are the principal sources for knowing current events, checking ads for employment, speaking and listening in the modern public square, and otherwise exploring the vast realms of human thought and knowledge”).

The Board's answer to this problem was to adopt IHRL as its framework. As it has since described it, "[a] defining theme of the Board's work is our conviction that Meta will make content moderation decisions in a fairer, more principled way if it bases them on the international human rights standards to which it has committed itself."²⁶⁶ The way the Board has applied IHRL is by cutting and pasting the three-part test under IHRL for state restrictions on speech into Board decisions, testing Meta's rules for legality, a legitimate aim, and necessity and proportionality.²⁶⁷

The Board's decision to invoke IHRL is not terribly surprising although it wasn't inevitable. The Board's founding documents direct it to pay particular attention to international human rights norms and principles and do not mention any other source of law, but they stop short of mandating that the Board apply IHRL directly.²⁶⁸ The ambivalence of the founding documents makes sense given, as I will discuss in detail shortly, it is not clear how IHRL should be applied in this context. There has nevertheless been a growing movement in academia and civil society for platforms to adopt IHRL as their normative framework.²⁶⁹ The Board may have been heeding these calls when it turned to IHRL. But it also may simply not have considered any alternatives.

Formally, the Board has justified its invocation of IHRL primarily by citing reports of a UN Special Rapporteur who endorsed this approach.²⁷⁰ The Board did not discuss the authoritative weight of such reports (which are persuasive but subsidiary authority under IHRL²⁷¹).²⁷² The Board has also cited Meta's voluntary commitment to respect human rights in accordance with the United Nations Guiding Principles on Business and Human Rights ("UNGPs")²⁷³ as a reason to

266. 2022 ANNUAL REPORT, *supra* note 21, at 51.

267. David Kaye (Special Rapporteur on Freedom of Expression and Opinion), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion & Expression*, at 20, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018).

268. *See supra* Section III.A.2.

269. *See Kaye, supra* note 267; *see also* Evelyn Douek, *The Limits of International Law in Content Moderation*, 6 U.C. IRVINE J. INT'L, TRANSNAT'L & COMPAR. L. 37, 38–39 (2021).

270. *See, e.g.*, Colombia Protests Case, 2021-010-FB-UA (Oversight Bd. Sept. 27, 2021), <https://www.oversightboard.com/decision/FB-E5M6QZGA/> [<https://perma.cc/3LXN-XVYL>] ("[T]he UN Special Rapporteur on freedom of opinion and expression has called on social media companies to ensure their content rules are guided by the requirements of Article 19").

271. *See* Statute of the International Court of Justice, art. 38(1)(d), June 26, 1945, 59 Stat. 1055, 1060.

272. *See* Andreas Kulick, *Meta's Oversight Board and Beyond – Corporations as Interpreters and Adjudicators of International Human Rights Norms*, 22 L. & PRAC. INT'L CTS. & TRIBS. 161, 179–80 (2022) (arguing that the Board fails to follow established interpretive methodologies).

273. *See* John Ruggie (Special Representative of the Secretary-General), *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, at 3, U.N. Doc. A/HRC/17/31 (Mar. 21, 2011) (describing UNGPs).

rely on the IHRL framework.²⁷⁴ In later decisions, the Board justifies its use of the IHRL framework self-referentially. In most recent cases, it provides as the reason to use the IHRL framework the fact that “[t]he Board has employed the three-part test based on Article 19 of the [International Covenant on Civil and Political Rights (“ICCPR”)] in all of its decisions to date.”²⁷⁵ This self-referential justification is now a ritual incantation in almost every decision, with the Board referring to its own prior decisions as “sources of authority,”²⁷⁶ which require the use of the IHRL analytical framework.

In other words, the Board has suggested that it is simple and self-evident that it should adopt the IHRL framework as the basis of all its decision-making. In fact, this framing belies much complexity. Contrary to what the Board’s cursory analysis might suggest, IHRL does not neatly map onto the problems content moderation raises.²⁷⁷ Crucially, IHRL, like First Amendment law, is a body of norms intended to constrain public authorities. It cannot simply be transposed from state-based jurisprudence and applied to the practices of private companies without interrogation of the very meaningful differences between these two contexts. Both private power and public power can threaten the speech rights of others, but the nature of the threat is different, and the normative framework for protecting speech in the two contexts must necessarily differ. Therefore, if IHRL is to be applied to content moderation by private companies, the key question is how such rules differ in the context of a private company versus when they are being applied to a state.

The characteristics of public and private speech regulators differ in several relevant and significant respects. What that means for their responsibilities is a deep and difficult question that is beyond the scope of this Article, but it is enough to note the contours of the problem and the obvious issues with which the Board is failing to engage. First, governments and platforms can threaten different sanctions. A government can lock you up, fine you, or change your legal status. Companies

274. *E.g.* Wampum Belt Case, 2021-012-FB-UA (Oversight Bd. Dec. 9, 2021), <https://www.oversightboard.com/decision/FB-L1LANIA7/> [<https://perma.cc/X9VU-T5NY>].

275. *Id.*

276. The Board first started doing this in the Knin Cartoon Case, 2022-001-FB-UA (Oversight Bd. June 17, 2022), <https://oversightboard.com/decision/FB-JRQ1XP2M/> [<https://perma.cc/3A3W-B86D>].

277. For more extensive analysis on this point, see Douek, *supra* note 269, at 39–41; Dvoskin, *supra* note 239, at 113–24; EMILY B. LAIDLAW, REGULATING SPEECH IN CYBERSPACE: GATEKEEPERS, HUMAN RIGHTS AND CORPORATE RESPONSIBILITY 196 (2015); Molly K. Land, *Regulating Private Harms Online: Content Regulation Under Human Rights Law*, in HUMAN RIGHTS IN THE AGE OF PLATFORMS 285, 292–94 (Rikke Frank Jørgensen ed., 2019); Susan Benesch, *But Facebook’s Not a Country: How to Interpret Human Rights Law for Social Media Companies*, 38 YALE J. ON REGUL. BULL. 86, 89–92 (2020); Rachel Griffin, *Rethinking Rights in Social Media Governance: Human Rights, Ideology and Inequality*, 2 EUR. L. OPEN 30 (2023).

cannot do these things. This does not mean their actions cannot have profound consequences, though — platforms can, in some cases, cut someone off from their livelihoods, their social circles, or prevent them from receiving everyday modern services like deliveries or cloud storage.²⁷⁸ Second, the social meaning of being sanctioned by a public actor and a private one may differ. State sanction usually carries with it a level of social stigma that may not attend sanction by a corporation. Third, while both governments and platforms may regulate speech to suit their own interests, their motives for doing so may (and likely in many cases do) differ. Free speech jurisprudence is especially vigilant about governments suppressing particular ideologies or opinions they dislike,²⁷⁹ but while companies may also have political agendas, the dominant concern is usually that the company's commercial interests will skew their decision-making. Fourth, states and companies have different kinds of expertise when it comes to assessing the harms and benefits of different kinds of speech restrictions. The government will be in a position to assess national security or foreign relations considerations, for example, that a private company simply will not. Conversely, a corporation might have more insight into the specific ways its products are exploited, such as manipulative influence operations or commercial scams, and greater capacity to intervene in a targeted way. Fifth, the costs of entry and exit for affected parties are different — Meta's platforms may be dominant, but it is still easier to set up an account on a different social networking service than it is to move countries.²⁸⁰ Sixth, the tools available to public and private speech regulators to enforce their rules are very different. Just as a matter of sheer technical capacity, platforms have far more ability to be flexible and nimble in their rule enforcement, and the online environment means they have more fine-grained control over and visibility into their speech environments than offline regulators.²⁸¹ Finally, and significantly, corporations have their own speech (or if you prefer, "business") rights and interests, which surely entitle them to some latitude a state should not have. But the question is: how much?

These important differences in how and why states and corporations regulate speech mean that developing a normative framework for

278. Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45, 57 (arguing that people use the Internet "not only to communicate, but also to buy and sell merchandise, find dates, make restaurant reservations, watch television, read books, stream music, and look for jobs").

279. See *Rosenberger v. Rector and Visitors of the Univ. of Va.*, 515 U.S. 819, 829 (1995) ("Viewpoint discrimination is thus an egregious form of content discrimination.").

280. See Rory Van Loo, *The Corporation as Courthouse*, 33 YALE J. ON REGUL. 547, 575 (2016).

281. See Evelyn Douek, *Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 792–93 (2021).

corporate content moderation raises new questions that cannot easily be answered by a preexisting normative framework designed for states — whether that normative framework be IHRL, or the First Amendment, or anything else. It is not an exaggeration to say that the central challenge of creating a normative foundation for platform content moderation is to account for the ways in which companies can regulate speech that a government cannot. If it were not for this difference, platforms should simply be required to observe the restrictions that are placed on government actors. That is, old precedents could just be applied directly. Of course, there would be difficult questions about how such precedents apply to new facts, as there always are, but there would be no deep theoretical knot to untangle. But a knot there is, and untangling it is the core challenge for the Board if it wants to create a body of principled content moderation rules. It cannot simply invoke IHRL and suggest that it readily provides the answers.

Instead of facing this challenge head-on, the Board's decisions only tend to confirm that the work that IHRL does in this context is largely symbolic.²⁸² The Board has barely discussed the differences between state and private speech regulators — it has not even discussed the ways in which the UNGPs explicitly distinguish a state's and a company's obligations under IHRL.²⁸³ Indeed, it more often emphasizes the way companies and states might face similar questions, rather than their differences.²⁸⁴ In no case has the use of IHRL meaningfully constrained the Board, nor could it. This is not a weakness of IHRL alone — because there is so little precedent about how to think about free speech interests in the context of private platforms, no body of preexisting norms answers the difficult questions for making rights-respecting content moderation decisions. This indeterminacy means that the Board could start from IHRL, Meta's values, the First Amendment, or first principles, and ultimately the utility of each of these frameworks would run out in the same way: on most questions, there is no preexisting body of precedent, and the Board will simply be left making a judgment call. But this indeterminacy is a weakness in IHRL that needs to be addressed if it is to be made into a meaningful normative framework for content moderation decisions.

282. Douek, *supra* note 269, at 56–58 (noting that IHRL will not provide determinate answers to most difficult content moderation questions and leave “considerable residual discretion in the interpretation and application” of itself); Griffin, *supra* note 277, at 50 (noting that, often, “[h]uman rights provisions are not doing much work here beyond providing general rhetorical support for policy arguments influenced by other political views about how online media should be governed”).

283. These principles distinguish states' “duty to protect” from businesses' “responsibility to respect” IHRL. See Ruggie, *supra* note 273, at iii; see Dvoskin, *supra* note 239, at 99, 103.

284. *Oversight Board Finds a Facebook's Rule Application Violates International Human Rights Law*, *supra* note 29, at 1972.

The Board initially acknowledged the centrality and importance of its role as the translator of existing norms to this new context. In one of its first decisions, it noted that the human rights obligations imposed on private companies are not the same as those imposed on governments, and it concluded accordingly that “clarifying the nature” of IHRL obligations on Meta was the “principal task facing this Board.”²⁸⁵ As Helfer and Land observe, “[T]he Board is uniquely placed to explain, through an accretion of case-specific decisions on content moderation and broader policy recommendations, how international rules that bind states apply to private social media companies.”²⁸⁶ And yet it has never really attempted that task. It has not clarified the nature of Meta’s obligations under IHRL at all. Instead, the conclusory way the Board adopted the IHRL framework is mirrored in the conclusory way it applies IHRL in individual cases.

From time to time the Board still notes in its decisions that the framework that applies to a private company will necessarily be different from that which applies to a government.²⁸⁷ But these comments tend to be cursory and are becoming less frequent. More often the Board does not advert to the difference at all. The Board instead regularly cites rules that apply to states and then proceeds to apply them to Meta, without comment, often using general and passive language like “restrictions on freedom of expression must [meet IHRL criteria],”²⁸⁸ eliding that the authorities to which they cite are exclusively state-based jurisprudence. In some cases, in what might be a Freudian slip, the Board gives incantations like “[t]he principle of legality under [IHRL] requires rules used by states to limit expression to be clear and

285. *Armenians in Azerbaijan Case*, 2020-003-FB-UA (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/FB-QBJDASCV> [<https://perma.cc/M8Q3-GC6W>].

286. Helfer & Land, *supra* note 249, at 2240.

287. *See, e.g.*, *Depiction of Zwarte Piet Case*, 2021-002-FB-UA (Oversight Bd. Apr. 13, 2021), <https://www.oversightboard.com/decision/FB-S6NRTDAJ> [<https://perma.cc/2MQK-6HQU>] (upholding a Facebook rule that the Board said would be impermissible for states to do under IHRL); *Swedish Journalist Reporting Sexual Violence Against Minors Case*, 2021-016-FB-FBR (Oversight Bd. Feb. 1, 2022), <https://www.oversightboard.com/decision/FB-P9PR9RSA> [<https://perma.cc/B6E2-627B>] (stating that “[a]lthough the ICCPR *does not create the same obligations for Meta as it does for states*, Meta has committed to respecting human rights” (emphasis added)); *Knin Cartoon Case*, 2022-001-FB-UA (Oversight Bd. June 17, 2022), <https://oversightboard.com/decision/FB-JRQ1XP2M/> [<https://perma.cc/3A3W-B86D>] (“The Board considered the factors in the Rabat Plan of Action . . . to guide its analysis, while accounting for the differences between international law obligations of states and human rights responsibilities of businesses.” (emphasis added)).

288. *See, e.g.*, *Shared Al Jazeera Post Case*, 2021-009-FB-UA (Oversight Bd. Sept. 14, 2021), <https://www.oversightboard.com/decision/FB-P93JPX02/> [<https://perma.cc/98VM-8ARS>]; *COVID Lockdowns in Brazil Case*, 2021-008-FB-FBR (Oversight Bd. Aug. 19, 2021), <https://oversightboard.com/decision/FB-B6NGYREK/> [<https://perma.cc/47V4-WGMH>].

accessible,”²⁸⁹ without noting that, well, Meta is not a state and so this rule does not technically apply. Far from working through the underlying normative questions, typically the Board conclusorily asserts that certain outcomes are supposedly required by IHRL but without explaining why.

A few concrete examples may be helpful. A particularly flagrant example of shallow reasoning is the *Russian Poem Case*.²⁹⁰ The case concerned a Facebook post comparing the Russian army in Ukraine to Nazis, and the post included a photo of what appeared to be a dead body, face down in the street.²⁹¹ The Board summarily held that Meta could not put a warning screen over the photo — a screen that any user could click though — because such a screen was not “necessary” under IHRL given that the photo did not show any “clear visual indicators of violence.”²⁹² No authority is cited for this proposition. No reasoning is offered. No justification is given for why IHRL would prevent a private company from giving users advance notice before they see what appeared to be a murder victim. No citation is given for the proposition because none exists — how warning screens should be used on private platforms is of course not a question IHRL authorities have considered before.

This illustrates one of the difficulties of applying IHRL in the context of content moderation: almost all issues are novel. How should Article 19 of the ICCPR apply to warning screens in social media newsfeeds?²⁹³ What about de-amplification of content? Or turning off certain account features, such as the ability to livestream, as sanction for breaking the rules? Prior IHRL authorities simply have not grappled with these questions. The Board therefore cannot rely on IHRL as is, without further explanation. It is not that the principles underlying IHRL cannot guide content moderation decision-making — the nature of common law reasoning is to take old cases and apply them to or distinguish them from new situations. But to do this requires work — work that the Board is just not doing in its decisions.

Conversely, there have been a few cases where the Board has found that Meta can depart from what IHRL would require of a state because it is a private company. But in these cases, too, the Board has failed to provide much explanation as to why. For example, the Board ruled that

289. Myanmar Bot Case, 2021-007-FB-UA (Oversight Bd. Aug. 11, 2021) (emphasis added), <https://oversightboard.com/decision/FB-ZWQUPZLZ/> [<https://perma.cc/ZF3X-2VVT>]; see also *COVID Lockdowns in Brazil Case*, 2021-008-FB-FBR (also silently noting ICCPR limitations on state restrictions on expression).

290. 2022-008-FB-UA (Oversight Bd. Nov. 16, 2022), <https://www.oversightboard.com/decision/FB-MBGOTVN8/> [<https://perma.cc/FEE4-XBFW>].

291. *Id.*

292. *Id.*

293. International Covenant on Civil and Political Rights art. 19, *adopted* Dec. 16, 1966, 999 U.N.T.S. 171.

Meta could remove content depicting blackface, including portrayals of Zwarte Piet, a Dutch cultural tradition, because such content creates a discriminatory environment for Black people, and this was so even though IHRL “would not allow a state to impose a general prohibition on blackface.”²⁹⁴ A majority of the Board found that because there was some evidence of severe harms at a societal and individual level stemming from Zwarte Piet and other stereotypes, “this justified Facebook adopting a policy that departs from the human rights standards binding states.”²⁹⁵ But the majority did not explain or justify this decision in any way. What exactly about the private status of Meta justified the departure from IHRL norms binding states in this case? Is it something to do with the different sanctions or social meaning of Meta’s rules? Is it something to do with the unique features of the online environment? Is this a one-off exception or some generalizable principle? The Board simply stated the outcome, leaving any deeper questions unanswered.²⁹⁶

The superficiality of the way the Board has adopted and applied the IHRL framework manifests in the Board’s reasoning more broadly. For example, despite adopting the norms and forms of judicial decision-making, the Board has generally been quite casual in its use and development of precedent. It invokes and purports to follow IHRL, but it has never discussed the facts of any IHRL decision. The Board’s use of its own decisions as precedent has also been perfunctory. It has never suggested anything other than that the application of a prior decision is clear and has ignored inconvenient tensions. In the policy advisory opinion in which the Board concluded that Meta should continue to remove COVID-19 misinformation, for example, the Board did not even mention that in one of its first decisions it had overturned Meta’s decision to take down a post containing misinformation because the harm it could cause was not sufficiently “imminent.”²⁹⁷

Or take three cases that raised the question of when Meta can or should add warning screens to violent content. In one case, the Board upheld Meta’s decision to restore a graphic video that depicted a wounded body in Sudan and apply a warning screen, stating that “[t]he warning screen does not place an undue burden on those who wish to see the content while informing others about the nature of the content and allowing them to decide whether to see it or not. The warning

294. Depiction of Zwarte Piet Case, 2021-002-FB-UA (Oversight Bd. Apr. 13, 2021), <https://www.oversightboard.com/decision/FB-S6NRTDAJ> [<https://perma.cc/2MQK-6HQU>].

295. *Id.*

296. Dvoskin, *supra* note 239, at 133 (similarly noting that “in the Black Pete case, the Board’s adoption of the IHRL project actually harmed the coherence and clarity of its decision”).

297. Claimed COVID Cure Case, 2020-006-FB-FBR (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/FB-XWJQBU9A> [<https://perma.cc/694C-4PTG>].

screen also adequately protects the dignity of the individual depicted and their family.”²⁹⁸ Five months later, in the *Russian Poem Case* described above, the Board overturned Meta’s decision to apply a warning screen to an image of a dead body in Ukraine, saying that “the use of a warning screen inhibits freedom of expression and is not a necessary response in this instance.”²⁹⁹ The latter case cited to the former case in its “sources of authority” but otherwise did not discuss it or attempt to resolve the inconsistencies.³⁰⁰ A minority of the Board disagreed with the majority, stating that “Meta may err on the side of prudence” in such cases, but the minority also did not invoke the prior decision.³⁰¹ It is not that the cases are necessarily inconsistent. The Board could have, for example, suggested that a difference in the level of graphic gore in the two cases led to its different conclusions. Instead, the Board simply did not discuss the prior case at all. In a third case, the Board changed tack again and upheld Meta’s decision to add a warning screen to a video that appeared to document war crimes committed by Azerbaijan in which Armenian prisoners of war and casualties were visible.³⁰² The Board relied on the *Sudan Graphic Video Case* and held that “providing users with the choice of whether to see disturbing content is a proportionate measure” that showed “respect for the rights of the prisoners and their families.”³⁰³ In this decision, the Board at least acknowledged the inconsistency with the *Russian Poem Case* but it did not explain in any detail how to reconcile the decisions, beyond perfunctorily suggesting that the content was less graphic in the earlier decision.³⁰⁴

By citing authorities but not meaningfully engaging with them, the Board gives the appearance of creating a body of precedent, without doing the hard work of reconciling cases that such a task entails. Some of the Board’s “doctrinal” incoherence, of which more examples could be given, no doubt stems from the differing make-up of different panels. But the Board has opted for unsigned opinions, and without knowing who makes decisions, the result is that the Board as a whole has the appearance of contradicting itself. It also means no individual member of the Board pays a reputational cost for poor work product.

A number of the Board’s practices have, intentionally or not, allowed this lack of deep thorough or coherent reasoning to fly under the

298. Sudan Graphic Video Case, 2022-002-FB-MR (Oversight Bd. June 13, 2022), <https://www.oversightboard.com/decision/FB-AP0NSBVC> [<https://perma.cc/L9XL-762T>].

299. Russian Poem Case, 2022-008-FB-UA Oversight Bd. Nov. 16, 2022), <https://www.oversightboard.com/decision/FB-MBGOTVN8/> [<https://perma.cc/FEE4-XBFW>].

300. *Id.*

301. *Id.*

302. Armenian Prisoners of War Video Case, 2023-004-FB-MR (Oversight Bd. June 13, 2023), <https://oversightboard.com/decision/FB-YLRV35WD/> [<https://perma.cc/Q3G4-2B6G>].

303. *Id.*

304. *Id.*

radar. First, the fact that it is deciding so few cases³⁰⁵ means the Board does not need to confront the task of reconciling its decisions often. This very fact illustrates why deciding more cases is important — it is through cases that the Board should be explaining its reasoning and how to weigh competing considerations in different contexts. An essential part of the Board’s role is to play a discursive role in the ongoing, ever-changing debate about online speech rules. It cannot do this through intermittent, inconsistent engagement.

Second, the shallowness of the Board’s reasoning is both less noticeable because of, and compounded by, the way the Board has decided to present minority opinions. Despite the complexity of the issues and the size of the Board, minority opinions are comparatively rare, with one being expressed in only fourteen out of thirty-nine cases.³⁰⁶ More importantly, the presentation of minority opinions is usually brief, low on detail, and does not note the size or composition of the minority. Thus, even when there is internal disagreement, the Board does not describe the nature of the disagreement in a way that helps readers understand the theoretical complexities of the cases.³⁰⁷ In these ways, the Board has opted for simplicity over depth and presented a unified front over ventilating points of contention. Perhaps this made sense in the Board’s early cases — providing the institution an opportunity to establish its authority as a cohesive institution rather than highlighting the differences of its members. But three years on, the lack of meaningful debate even among Board members only serves to highlight the overall shallowness of the opinions.

Third, the Board’s decisions cannot be appealed, do not have to be interpreted by lower courts, cannot be invoked by litigants seeking to make their own cases, and are generally not embedded in a broader community of stakeholders that carefully read the Board’s jurisprudence as whole. For all the attention that the Board gets, little of it engages with the Board’s work at a level of detail that would highlight these kinds of inconsistencies or tensions. As a result, generally speaking, the Board’s reasoning and modes of analysis remain relatively unexamined.

The above critiques of the Board’s lack of normative framework or doctrinal coherence are not mere academic nitpicking or law review quarterbacking. Superficial reasoning undermines the guidance the Board’s decisions give to Meta in future cases (pity the poor content

305. See *supra* Section III.A.1.

306. As of July 30, 2023. For an example dissenting opinion, see *Russian Poem Case*, 2022-008-FB-UA (“A minority of the Board finds that the warning screen was a necessary and proportionate measure that was appropriately tailored to encourage participation and freedom of expression.”).

307. See also Helfer & Land, *supra* note 130, at 2284 (similarly noting that “[a]llowing individual Board members to write separate dissenting or concurring opinions will help identify areas where norms are in flux and would promote dialogue about those junctures”).

moderator who must determine when applying a warning screen violates IHRL in the Board's eyes, let alone why!), to users about the rules that apply to them, and to regulators for understanding these complex systems. It makes engaging with the Board's decisions harder when the bases for those decisions are obscure. And it also undermines the justification for the Board itself if it fails to engage with one of the most important tasks it was set up to perform — starting to provide a common language and framework for thinking about how to justify platform rules that are different from public free speech laws in a coherent and thoughtful way. Instead, the Board's reasons have all the trappings of jurisprudence, but too often they lack the underlying substance.

2. Formalistic Measures of Impact

If the Board's guiding normative framework is unclear, the criteria for assessing the Board's material impacts — that is, the effects its work has in the world — are even more so. What impact the Board's decisions have had on the world directly is almost impossible to say, and the Board itself has not prioritized answering this question.

It is somewhat striking how little material impact comes up in the discourse around the Board. There will often be a flurry of commentary and news articles when the Board hands down a case,³⁰⁸ but little (if any) reporting on Meta's responses to, or implementation of, the Board's recommendations. But the Board's decisions are the least consequential part of the entire process in a material sense, especially because its recommendations are nonbinding. All that matters, if the Board is to have more than rhetorical impact, is what Meta does in response. And yet it is the moment of decision that draws public attention. This dynamic is not unique to the Board and illustrates a deeper problem in how content moderation governance is evaluated. Anecdotes about individual cases and decisions capture the imagination far better than questions about system design.³⁰⁹ The story of the Board overturning, rebuking, and putting Meta in its place is far more attention-grabbing than, say, whether Meta publishes error rates for its Media Matching Service bank of violating images.³¹⁰ But in substance, the latter matters far more for systemic improvement. And what happens as a result of the publication of those error rates (that is, does it lead to some other form of accountability or reform?) matters even more still.

There are a few problems with the way the Board appears to think about impact. The first relates to the Board's emphasis on ever more

308. See King, *supra* note 11; Nix, *supra* note 12; Paul, *supra* note 13; Hern, *supra* note 14; Cabato & Tan, *supra* note 15.

309. See Douek, *supra* note 127, at 559–60.

310. Colombian Police Cartoon Case, 2022-004-FB-UA (Oversight Bd. Sept. 15, 2022), <https://www.oversightboard.com/decision/FB-I964KKM6> [<https://perma.cc/VC7C-FCPP>].

procedure as the indicia of effective governance. The Board's recommendations often reflect an assumption that giving users more detailed rules, more extensive reasons for decisions, more opportunities to appeal, and in general more due process, will necessarily result in a better and more legitimate system overall. But this assumption reflects a legalistic procedural fetishism that depends on abstract notions of legitimacy and accountability not always borne out in practice, and it does not acknowledge the costs of piling on procedural requirement after procedural requirement.³¹¹ The Board's intuitions that Meta's content moderation needs more individual due process cannot be merely assumed, and the Board should treat this as a hypothesis to be subject to testing and revision. Instead, the Board's interpretation of IHRL has been extremely formalistic and unconcerned with practical impacts. For example, the Board's interpretation of the principle of "legality" under IHRL that requires "rules that limit expression to be clear and publicly accessible" has resulted in many recommendations that Meta add additional wording to its already voluminous Community Standards.³¹² But the Board has not questioned whether this actually results in better outcomes for any users or the public more broadly.

Another category of problems arises as a result of the difficulty of quantifying impact. The Board does track the results of its recommendations, but in a narrow way. The Board tracks Meta's implementation of its recommendations, but not the effects of that implementation. As evidence of its importance, the Board notes that "Meta has publicly recognized how our recommendations are changing its behavior."³¹³ Tracking implementation alone, however, shows the Board's impact on Meta, not the Board's impact on the world. When Meta adds definitions to "non-medical drugs" and "pharmaceutical drugs" to its Community Standards or starts giving users more detailed notifications for specific policy violations for hate speech,³¹⁴ does that actually have meaningful effects on how speech flows through its platforms, let alone society more broadly? And . . . are they good effects? While the Board's impact on Meta may be relevant to the question of the Board's impact on

311. Nicholas Bagley, *The Procedure Fetish*, 118 MICH. L. REV. 345, 369 (2019); see also Douek, *supra* note 127, at 577–83; Elettra Bietti, *A Genealogy of Digital Platform Regulation*, 7 GEO. L. TECH. REV. 1, 45 (2023) ("Governance efforts and law-like procedures . . . become a façade with little intrinsic value and detrimental effects for the public.").

312. Violence Against Women Case, 2023-002-IG-UA, 2023-005-IG-UA (Oversight Bd. July 12, 2023), [https://www.oversightboard.com/decision/IG-H3138H6S/\[https://perma.cc/6DL2-F57H\]](https://www.oversightboard.com/decision/IG-H3138H6S/[https://perma.cc/6DL2-F57H]); see, e.g., York & Kayyali, *supra* note 258 ("Out of the Oversight Board's 18 recommendations to date, nearly every one of them echoes the Santa Clara Principles' recommendation to disclose 'detailed guidance to the community about what content is prohibited.'").

313. 2022 ANNUAL REPORT, *supra* note 21, at 19.

314. OVERSIGHT BD., OVERSIGHT BOARD Q3 2022 TRANSPARENCY REPORT 21 (Dec. 2022), <https://oversightboard.com/news/1147590722555454-oversight-board-publishes-transparency-report-for-third-quarter-of-2022/> [<https://perma.cc/6FH9-UY72>].

the wider world, it is not a perfect or even necessarily reliable proxy. Without more, knowing only that Meta made some change in its practices is not evidence of positive material impact.

Even though the Board clearly wants to tout its positive impacts, the Board is not incentivized to measure its material impact in this way for three reasons: first, broader material impact is hard to measure; second, there is no normative agreement on what impact is positive, making outcome goals difficult to define; third, it is easier for the Board to sell a story of success if it focuses on the narrower question of its impact on Meta than some broader question about its effects in the world.

First, the challenges of measuring the impact of various speech rules and governance practices are substantial. Determining even the first-order effects of speech decisions is hard, let alone the flow-on effects of any changes in a complex system. Every system of speech governance has struggled with it. Take the example of hate speech — perhaps one of the most controversial free speech issues, and one in which you might think that there is a perfect natural experiment: the United States is unique in its almost blanket protection of hate speech.³¹⁵ Research has not been able to settle long-standing debates about the relationship between hate speech laws and hate speech, or the effects of hate speech itself.³¹⁶

But such difficulty is not sufficient reason to ignore the importance of substantive social impact as a metric of good governance. The consequences of rules and their enforcement have for society at large “are the broadest and most diffuse of all effects, but in the final analysis, they are the most important.”³¹⁷ The whole project of speech governance rests on an assumption that institutions matter because they affect the world. The Board’s existence and decisions assume that the ways in which speech rules are written and enforced matters in some material way. Perhaps the best available measures of impact will always be indeterminate, disputed, and contingent, but they should not be neglected for this reason alone. Working out the substantive impact of various decisions requires at least starting with the right questions. What should matter to the Board is not only whether Meta implements its

315. Michel Rosenfeld, *Hate Speech in Constitutional Jurisprudence: A Comparative Analysis*, 24 *CARDOZO L. REV.* 1523, 1523 (2003); Frederick Schauer, *The Exceptional First Amendment*, in *AMERICAN EXCEPTIONALISM AND HUMAN RIGHTS* 29, 35 (Michael Ignatieff ed., 2009).

316. Compare Katharine Gelber, *Hate Speech — Definitions & Empirical Evidence*, 32 *CONST. COMMENT.* 619, 623 (2017) (compiling evidence of harms of hate speech), and James Hawdon, Atte Oksanen & Pekka Räsänen, *Exposure to Online Hate in Four Nations: A Cross-National Consideration*, 38 *DEVIANANT BEHAV.* 254, 262 (2017) (suggesting a causal link between hate speech laws and exposure to online hate), with NADINE STROSSEN, *HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP* 121–56 (compiling studies questioning the harm of hate speech and the effectiveness of laws against it).

317. James P. Levine, *Methodological Concerns in Studying Supreme Court Efficacy*, 4 *LAW & SOC’Y REV.* 583, 586 (1970).

recommendation to give users better notifications about hate speech violations, for example, but also whether better notifications in some way benefit not only users but the speech environment more broadly. Focusing on the former at the expense of the latter is goal displacement, with the Board becoming focused on what is measurable rather than what is impactful.³¹⁸ The disruption of the platform era offers an opportunity to start answering questions about speech governance that have typically been answered by intuition alone. Because much speech online is tractable and searchable, rather than ephemeral, there is now more data available than ever before about the way speech flows through society.³¹⁹

The task of measuring impact is made even harder by ambiguity in defining the Board's relevant constituency. The first question must be: who exactly is the Board serving? Is it Meta's users, or society more generally? Zuckerberg seemed to envision the former, writing that "[j]ust as our board of directors is accountable to our shareholders, [the Board] would be focused only on our community."³²⁰ The Board also suggests that it "was established . . . to promote the rights and interests of users."³²¹ And many of its recommendations focus on Meta's due process obligations to the people directly affected by decisions rather than the diffuse impact Meta makes in society. But users' interests will not always align with the broader public interest. And the Board's current emphasis on users' interests and rights — although perhaps an attractively more modest domain — is somewhat in tension with the UNGPs it purports to apply, which emphasize business enterprises' responsibility to address any adverse human rights impacts.³²² A proper application of these principles thus requires the Board to be attentive to Meta's (and its own) impacts on society beyond Meta's users.³²³ The Board needs to be clearer about its community of interest in order to determine if the Board is serving it well.

But this leads to the second problem, because evaluating (not just measuring) performance on the basis of real-world impact requires some notion of the normative good that the Board is trying to achieve.

318. John Bohte & Kenneth J. Meier, *Goal Displacement: Assessing the Motivation for Organizational Cheating*, 60 PUB. ADMIN. REV. 173, 174 (2000).

319. Nathaniel Persily & Joshua A. Tucker, *Introduction*, in SOCIAL MEDIA AND DEMOCRACY 1, 8 (Nathaniel Persily & Joshua A. Tucker eds., 2020) ("For the first time in human history, we have real time records of millions – if not billions – of people as they discuss politics, share information about politics, and organize politically. Each of these actions simultaneously produces an archived, digitized record.").

320. Zuckerberg, *supra* note 49.

321. 2021 ANNUAL REPORT, *supra* note 112, at 3 (emphasis added).

322. Ruggie, *supra* note 273, at 13.

323. See *Out of "Site," Out of Mind: The Facebook Oversight Board's Exclusion of Non-User Rightsholders*, U.C. IRVINE SCH. OF LAW INT'L JUST. CLINIC (Mar. 5, 2021), <https://ijclinic.law.uci.edu/2021/03/05/out-of-site-out-of-mind/> [https://perma.cc/ZF2N-DXEU].

Alas, defining the normative goals of speech governance is even more fraught than trying to measure its impact.

Recall, for example, Feldman’s goal that the Board should be “an effective counterweight to censorship.”³²⁴ On this view, the measure of success is whether the Board has decreased the amount of content that Meta takes down. This is a metric that can be measured. And, indeed, it’s possible that on this metric the Board has been quite successful. In all but three³²⁵ of the cases where the Board overturned Meta (that is, changed the substantive outcome), the Board ordered a post to be reinstated that had initially been taken down. In fact, in its entire corpus of cases, the Board has only thought that a post should be taken down eight times.³²⁶ It seems, then, that the Board itself views its role in the same vein as Feldman — to push back on Meta’s removals. If this is your conception of the substantive good, then the Board may be “working.” But of course, hardly everyone agrees that the problem with content moderation is that platforms take down too many posts. Many believe the main problem with Meta’s content moderation is that it leaves far too much harmful content up.³²⁷

Whatever the Board’s own goals for its substantive impact, it has not been explicit about them, let alone measured progress towards them. Instead, the Board has shown relative disinterest in this project³²⁸ or suggested that Meta is withholding relevant information.³²⁹ For example, one of the key avenues of potential impact of the Board is Meta’s obligation to implement the Board’s decisions with respect to

324. Feldman, *supra* note 69, at 102.

325. Knin Cartoon Case, 2022-001-FB-UA (Oversight Bd. June 17, 2022), <https://oversightboard.com/decision/FB-JRQ1XP2M/> [<https://perma.cc/3A3W-B86D>]; Brazilian General’s Speech Case, 2023-001-FB-UA (Oversight Bd. June 22, 2023), <https://www.oversightboard.com/decision/FB-659EAWI8/> [<https://perma.cc/79HA-AFE5>]; Cambodian Prime Minister Case, 2023-003-FB-MR (Oversight Bd. June 29, 2023), <https://www.oversightboard.com/decision/FB-6OKJPNS3/> [<https://perma.cc/38PC-P6WH>].

326. Not including the *Trump Suspension Case*, *supra* note 1; *Knin Cartoon Case*, 2022-001-FB-UA; *Armenians in Azerbaijan Case*, 2020-003-FB-UA (Oversight Bd. Jan. 28, 2021), <https://www.oversightboard.com/decision/FB-QBJDASCV> [<https://perma.cc/M8Q3-GC6W>]; *Zwarte Piet Case*, 2021-002-FB-UA (Oversight Bd. Apr. 13, 2021), <https://www.oversightboard.com/decision/FB-S6NRTDAJ> [<https://perma.cc/2MQK-6HQU>]; *Tigray Commc’n Affs. Bureau Case*, 2022-006-FB-MR (Oversight Bd. Oct. 4, 2022), <https://www.oversightboard.com/decision/FB-E1154YLY/> [<https://perma.cc/RN3L-TCCN>].

327. See, e.g., Andrew Marantz, *Why Facebook Can’t Fix Itself*, *NEW YORKER* (Oct. 9, 2020), <https://www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself> [<https://perma.cc/HC4B-HRX2>].

328. In the *Brazilian General’s Speech Case*, the Board suggested for the first time that it would “pay special attention to Meta’s application of its decision to identical content with parallel context that has remained on the company’s platforms,” but did not specify how it would do so. 2023-001-FB-UA.

329. Naomi Shiffman, Carly Miller, Manuel Parra Yagnam & Claudia Flores-Saviaga, *Commentary, Burden of Proof: Lessons Learned for Regulators from the Oversight Board’s Implementation Work*, *J. ONLINE TR. & SAFETY*, Feb. 2024, at 7, <https://doi.org/10.54501/jots.v2i2.168> [<https://perma.cc/U852-XTWG>].

individual pieces of content to “identical content with parallel context.”³³⁰ But neither the Board nor Meta have publicly tracked how broadly Meta has carried out this obligation, and thus even the direct reach of the Board’s decisions remains unknown. Nor has the Board sought feedback or evidence about the impact of its many procedural recommendations, like requiring Meta to add more language to its Community Standards or provide more reasoning to users. For the Board to establish meaningful systemic impact on Meta’s content moderation, it needs to track far more than whether Meta has restored the few dozen pieces of content that the Board has examined directly, or whether Meta has ticked the box on procedural reforms. The Board may be aware of this weakness in its reporting and has started to talk more about evaluating not only the implementation of a recommendation but also its impact.³³¹ In the transparency report for the first quarter of 2023, the Board included additional data about the impacts of two prior recommendations for the first time, stating that Meta had shared data showing that more content was being sent to human review than would have otherwise as a result of the Board’s recommendations.³³² But this is a great illustration of exactly my critique — the mere fact that more content was sent for human review cannot, without more, merely be assumed to be beneficial unless your definition of benefits revolves entirely around proceduralism.

The final reason the Board may not be tracking its broader impact is more cynical: the results may be less impressive compared with the Board’s direct impact on Meta alone. Being able to show that Meta took X number of concrete steps in response to its rulings helps the Board establish its credentials in its role as watchdog and is far less indeterminate and risky than trying to measure downstream impacts.

3. Persistent Strategic Avoidance

Surprisingly, for an institution set up to “help [Meta] resolve . . . what to take down, what to leave up and why,”³³³ the Board has a clear pattern of ducking the most controversial questions that come before it. In the most difficult cases, the Board is very eager to tell Meta it did something wrong but much less eager to tell Meta how to get things right. This is a much more risk averse approach: Meta cannot ignore, and people cannot disagree with, a decision if no decision is made. But

330. CHARTER, *supra* note 51, art. 4 (“In instances where Facebook identifies that identical content with parallel context — which the board has already decided upon — remains on Facebook, it will take action by analyzing whether it is technically and operationally feasible to apply the board’s decision to that content as well.”); *see also* BYLAWS, *supra* note 51, art. 2, § 2.3.1 (“Implementation of Board Decisions”).

331. Shiffman et al., *supra* note 329, at 7.

332. Q1 2023 QUARTERLY TRANSPARENCY REPORT, *supra* note 133, at 17.

333. *The Oversight Board*, *supra* note 65; OVERSIGHT BD., *supra* note 65.

the persistence with which the Board does this undermines its utility as a decision-maker of last resort that offers some measure of finality in content moderation debates.

Perhaps the clearest example was the Board's decision with respect to the deplatforming of then-President Trump.³³⁴ Instead of answering the question of whether Trump's account on Facebook should be reinstated, the Board instead rebuked Meta for the way in which it had made the decision and sent it back to Meta, giving Meta six months to make a new decision.³³⁵ By avoiding making the hard call itself, "[t]he realist take is that the [Board] was trying to avoid controversy and prioritize its own legitimacy."³³⁶ In a sense, this was a politically deft move. A non-decision with respect to Trump's account forestalled the public outrage that might have followed a substantive decision either way. It also sidestepped being too confrontational with Meta. By slapping Meta over the knuckles for procedural deficits in its decision about Trump's account, the Board looked like it was holding Meta accountable without actually testing the limits of its coercive authority — Meta could not ignore or contradict the Board's decision if the Board didn't make any decision. But while politically deft, the decision was also an abdication of duty. Making this kind of hard decision was squarely within the Board's job description to "exercise independent judgment over some of the most difficult and significant content decisions."³³⁷ The Board ducked its responsibility in order to avoid appearing weak or becoming unpopular.

At the time, the decision was analogized to *Marbury v. Madison*³³⁸ because it was the first big display of the Board's power.³³⁹ The analogy is apt, but not for that reason. By contrast with the Supreme Court's powers of judicial review, it was never in doubt that the Board had the formal power to issue a decision in the *Trump Suspension Case* — indeed, such decisions were the reason it was created. But the analogy fits as an example of judicial statecraft in which "the Court protected

334. *Trump Suspension Case*, *supra* note 1.

335. *Id.*

336. Evelyn Douek, *It's Not Over: The Oversight Board's Trump Decision Is Just the Start*, LAWFARE (May 5, 2021), <https://www.lawfareblog.com/its-not-over-oversight-boards-trump-decision-just-start> [<https://perma.cc/BV9N-X69B>].

337. Clegg, *supra* note 50 ("[W]e have created and empowered a new group to exercise independent judgment over some of the most difficult and significant content decisions.").

338. 5 U.S. (1 Cranch) 137 (1803).

339. See, e.g., Jeff Neal, *Did Facebook's Oversight Board Get the Trump Decision Right?*, HARV. L. TODAY (May 5, 2021), <https://hls.harvard.edu/today/did-facebooks-oversight-board-get-the-trump-decision-right/> [<https://perma.cc/AMZ7-MK3W>] ("The decision nevertheless tried to establish a couple of important principles, including most fundamentally the power of the board to say that Facebook got it wrong. So, in that sense, that is something that the Court did in *Marbury*."); Nilay Patel, *Facebook's Oversight Board Has Upheld Trump's Ban — What's Next?*, THE VERGE (May 11, 2021), <https://www.theverge.com/22428768/facebook-oversight-board-trump-ban-kate-klonick-interview> [<https://perma.cc/8VWX-76L9>].

itself from immediate political attack by finding that under the particulars of the case it was without jurisdiction to act.”³⁴⁰

This was merely the first example of what would become a regular practice of the Board sending the most high-profile and difficult decisions back to Meta settle for the company to decide. The Board’s decision about COVID-19 misinformation is another example.³⁴¹ In July 2022, Meta asked the Board for advice about “whether Meta’s current measures to address COVID-19 misinformation under [its] harmful health misinformation policy continue to be appropriate” in light of the changing pandemic conditions.³⁴² Nearly a year later, in April 2023, the Board issued its opinion — but not an answer. Its advice was that Meta should maintain its current policy, but Meta should also “begin a process to reassess each of the 80 claims it currently removes, engaging a broader set of stakeholders.”³⁴³ To which one might respond: wasn’t that exactly what Meta was asking the Board to do? It’s not hard to guess why the Board might not want to give a definitive answer, though. How platforms have dealt with COVID-19 misinformation has been one of the most politically controversial topics in content moderation.³⁴⁴ By again rebuking Meta for inadequate procedures and consultation about its policies but not providing a definitive answer on what the policy should be, the Board was able to perform toughness without wading into a political quagmire.

This dynamic is not confined to the most high-profile cases. The Board took the same approach in a perhaps less obviously consequential and lower-profile context — Meta’s policies around nipples. How Meta should treat nipples on its platforms is one of the earliest and most enduring content moderation controversies.³⁴⁵ It raises (perhaps surprisingly) difficult normative questions. From its earliest days, Meta made a business decision to generally ban adult nudity.³⁴⁶ Such a rule aligns with many users’ preferences and prevents Meta’s platforms

340. Rosalind Dixon & Samuel Issacharoff, *Living to Fight Another Day: Judicial Deference in Defense of Democracy*, 2016 WIS. L. REV. 683, 686.

341. Removal of COVID-19 Misinformation, PAO-2022-01 (Oversight Bd. Apr. 20, 2023), <https://www.oversightboard.com/decision/PAO-SABU4P2S/> [<https://perma.cc/2NW3-RD9G>].

342. Nick Clegg, *Meta Asks Oversight Board to Advise on COVID-19 Misinformation Policies*, META NEWSROOM (July 26, 2022), <https://about.fb.com/news/2022/07/oversight-board-advise-covid-19-misinformation-measures/> [<https://perma.cc/YF45-QASM>].

343. *Removal of COVID-19 Misinformation*, *supra* note 341.

344. See, e.g., Will Oremus, *How Social Media ‘Censorship’ Became a Front Line in the Culture War*, WASH. POST (Oct. 28, 2022), <https://www.washingtonpost.com/technology/2022/10/09/social-media-content-moderation/> [<https://perma.cc/GH69-UF5L>].

345. Soraya Chemaly, *#FreeTheNipple: Facebook Changes Breastfeeding Mothers Photo Policy*, HUFFPOST, https://www.huffpost.com/entry/freethenipple-facebook-changes_b_5473467 [<https://perma.cc/37JX-JABZ>].

346. Jessica Mintz, *Facebook Nudity Policy Draws Nursing Moms’ Ire*, ASSOCIATED PRESS (Jan. 1, 2009), <https://www.sandiegouniontribune.com/sdut-tec-facebook-breast-feed-ing-010109-2009jan01-story.html> [<https://perma.cc/33C6-H8CF>].

from becoming flooded with porn. As a business decision this might be a simple call, but as a speech decision it is far more complicated. A ban on female-presenting nipples perpetuates certain social norms about sexualization that many disagree with, stigmatizes practices such as breastfeeding, and has disparate impact on already marginalized communities such as women, societies that traditionally go bare-chested, and LGBTQ individuals. A broad-based ban on adult nudity of the kind Meta has historically had could never be imposed by a state in accordance with IHRL, or under most national constitutions. The question of how to treat nipples therefore raises thorny questions about how to reconcile a platform's right to define what kind of product experience it wants to provide its users in accordance with its business interests, on the one hand, with traditional free speech norms, on the other. But the Board punted this decision too when it came up in the *Gender Identity and Nudity Cases*.³⁴⁷ Rather than recommend a policy to Meta, it simply said that Meta should conduct a human rights impact assessment and then “define clear, objective, rights-respecting criteria to govern the entirety of its Adult Nudity and Sexual Activity policy . . . consistent with international human rights standards.”³⁴⁸ But surely, defining such criteria was a core aspect of the Board's intended role.

The same political dynamics may also partially explain the Board's low caseload. Fewer cases mean fewer opportunities for confrontation or mistake. It is notable how rarely the Board has attracted controversy, given how many controversial cases it no doubt could have selected. As the Board itself emphasizes, it receives many, many appeals, and presumably could take cases on almost any topic. But the Board never has to take a case it doesn't want, and it has avoided touching many of content moderation's livest wires.

From the perspective of evaluating the Board as a governance institution, this pattern seems like a failure to meet its responsibility. But as a matter of politics, the Board's desire to avoid controversy and sidestep hard calls is rational. The Board is not actually incentivized to weigh in on intractable normative and moral debates, despite this being its ostensible purpose; instead, the Board is incentivized to perform the role of a solemn governance institution, without upsetting too many people. It should therefore not be surprising that the Board has focused so much on the appearance of its judgments and its public communications, and less on its substantive productivity and impact. It is these incentives to which the next Part turns.

347. *Gender Identity and Nudity Cases*, 2022-009-IG-UA and 2022-010-IG-UA (Oversight Bd. Jan. 17, 2023), <https://www.oversightboard.com/decision/BUN-IH313ZHJ/> [<https://perma.cc/ZEP2-87KA>].

348. *Id.*

IV. LEGITIMACY'S EMPTY PROMISE

The story of the Board's success is therefore complicated and, this Part argues, ultimately suggests that the pursuit of sociological legitimacy for its own sake creates perverse incentives. Legitimacy lacks substantive content — in some ways, that's the very point. Agreeing on the substantive goals of content moderation is hard; agreeing we should have institutions worthy of respect is less so. But this breeds complacency about what governing institutions can and should achieve. The Board's focus on legitimacy building has come at the cost of a more ambitious, and more substantive, vision for content moderation governance.

The positive version of the story is that the Board has attained far more public attention and acceptance than many expected. For Meta and the Board, this is an indisputable win. Meta's creation of the Board and outsourcing of certain kinds of authority evinces Meta's belief that reestablishing public trust was more important to its business than any particular content moderation decision or rule. That is, that legitimacy matters and has independent value to Meta. The subsidiary goals for the Board described in Part II were all instrumental to this central aim.

But for the rest of us, the sociological legitimacy of the Board — indeed, of any platform governance institution — should theoretically depend on the achievement of other ends, not be an end in and of itself. And here, the story is more complicated. The previous Part showed that while the Board has unquestionably brought some benefits, it has also neglected core aspects of its role.

Why has the Board neglected these tasks? It seems hard to make sense of. Answering difficult normative questions was the reason the Board was created, and the Board was designed with this task in mind. The Board is not made up of data scientists or engineers, and it does not perform ongoing audits of Meta's systems. It comprises scholars, judges, human rights activists — that is, the kinds of people who are better placed to opine on the normative problems of content moderation, rather than its technical challenges. And yet, the Board has eschewed this kind of theory almost entirely. At the same time, the Board has hardly forsaken theory in the interests of a more practical and empirically grounded approach. The Board does not define, let alone measure, its goals in concrete terms. And despite all the big talk about “help[ing] Meta answer some of the most difficult questions around freedom of expression online,”³⁴⁹ often the Board dodges clear answers in the most fraught cases. This Part argues that these choices — initially baffling — make sense once the Board is understood as a political actor responding to its political environment.

349. *The Oversight Board*, *supra* note 65.

Understanding this political environment is important for debates far beyond Meta and the Board. The Board of course matters on its own terms. Billions of people use Meta’s products, and the company makes millions of speech decisions every hour.³⁵⁰ A mechanism to bring transparency and accountability to this expansive global system of speech regulation is presumptively important. And this is all the more true given Meta and the Board’s ambitions to grow the Board’s oversight and influence. Board members have from its earliest days said they hope the project can expand and ultimately include reviewing content moderation decisions from platforms beyond Meta’s.³⁵¹ Professor Thomas Kadri has carefully documented the ways Meta and the Board have made clear their more expansive vision for the institution.³⁵² Both have pushed for the Board to become a frame of reference for excellence in content moderation governance across the industry.³⁵³ Mark Zuckerberg explicitly pointed to the Board as an example of the kind of regulation that lawmakers should be thinking about,³⁵⁴ and Meta representatives have mentioned it in Congressional hearings.³⁵⁵ Meta’s website proclaims the hope that “the [B]oard can serve as a model for the future of content governance across our industry.”³⁵⁶ The Board has made it clear that it is “interested in working with companies that share our belief that transparent and accountable content governance, overseen by independent bodies, is an essential part of creating” a rights-respecting online public sphere.³⁵⁷ Clearly, the Board hopes to directly impact far more than merely Meta’s services.

350. See *supra* note 124 and accompanying text.

351. See, e.g., Taylor Hatmaker, *Facebook Oversight Board Says Other Social Networks ‘Welcome to Join’ if Project Succeeds*, TECHCRUNCH (Feb. 11, 2021), <https://techcrunch.com/2021/02/11/facebook-oversight-board-other-social-networks-beyond-facebook/> [<https://perma.cc/LZ6C-V6VD>]; @OversightBoard, X (FORMERLY KNOWN AS TWITTER) (Nov. 16, 2022), <https://twitter.com/OversightBoard/status/1592814589414367233> [<https://perma.cc/WPR7-ZR5T>].

352. Kadri, *supra* note 44, at 176–85.

353. *Id.*

354. Mark Zuckerberg, *The Internet Needs New Rules. Let’s Start in These Four Areas.*, WASH. POST (Mar. 30, 2019), https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html [<https://perma.cc/5EJV-29HT>].

355. See, e.g., *Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our Discourse and Our Minds Before the Subcomm. on Priv., Tech., & L. of the S. Comm. on the Judiciary*, 117th Cong. 4 (2021) (statement of Monika Bickert, Vice President for Content Policy, Facebook) (discussing the progress made by the Oversight Board); see also Kadri, *supra* note 44, at 195; cf. *Disinformation Nation: Social Media’s Role in Promoting Extremism and Misinformation Before the Subcomm. on Comm’n and Tech. and the Subcomm. on Consumer Prot. & Com. of the H. Comm. on Energy & Com.*, 117th Cong. 7 (2021) (statement of Mark Zuckerberg, Chief Exec. Officer, Facebook) (suggesting third party bodies as a source of standards).

356. *Creating the Oversight Board*, META TRANSPARENCY CTR. (Jan. 19, 2022), <https://transparency.fb.com/oversight/creation-of-oversight-board/> [<https://perma.cc/KTM7-WPGS>].

357. 2022 ANNUAL REPORT, *supra* note 21, at 63.

But, perhaps more fundamentally, the lessons of the Board experiment matter because debates around the Board provide a kind of microcosm through which broader debates about content moderation can be examined. The increasing public, academic, and civil society acceptance of and engagement with the Board reveals something important about the broader political environment in which the Board operates. And importantly, the political incentives to which the Board is responding will also exist for any content moderation governance institution. What the Board experiment shows is that “legitimacy” can be won through the presence of formalistic indicia including the appearance of reasoning, consultation, or certain other elements of due process. These indicia are seen as intrinsically beneficial, even without any further inquiry into their effects, and they are increasingly being mandated across the industry.

This juridical and legalistic model of governance that the Board exemplifies in particular is a tool increasingly favored by regulators.³⁵⁸ As Rachel Griffin has argued, the Board’s work and much regulation, especially in Europe, reflect a line of thought that believes in the legitimating power of rule of law procedural norms.³⁵⁹ Yoel Roth has observed the same phenomenon, bemoaning the “ever-more-formalized set of requirements around platform conduct” that tech companies and their regulators are turning to in order to try to build trust.³⁶⁰

Indeed, legislatures are writing and enacting laws that would require of all platforms the kinds of things that the Board’s recommendations to Meta seek to get Meta to adopt. For example, the European Union’s Digital Services Act (“DSA”) — set to become one of the most globally influential pieces of legislation regarding content moderation — shares many of the same goals as the Board’s recommendations.³⁶¹ It includes requirements for extensive procedural protections in every case: platforms need to provide reasons for any content removal,³⁶² a right of appeal open for six months in all cases,³⁶³ a human in the loop for all appeals,³⁶⁴ and a further right of appeal to a third-party arbitrator.³⁶⁵ The Board itself has expressed an interest in becoming a formal part of this regime.³⁶⁶ Laws in Texas and Florida similarly

358. See Douek, *supra* note 127, at 544.

359. Griffin, *supra* note 227, at 55–58.

360. Yoel Roth, *Content Moderation’s Legalism Problem*, LAWFARE (July 24, 2023), <https://www.lawfaremedia.org/article/content-moderation-s-legalism-problem> [<https://perma.cc/VQ9B-6AE3>].

361. See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1.

362. *Id.* art. 17(1)(a).

363. *Id.* art. 20(1).

364. *Id.* art. 20(6).

365. *Id.* art. 21(1).

366. Kadri, *supra* note 44, at 194; Levy, *supra* note 171.

require individual notice, appeal, and reasons.³⁶⁷ The Irish government commissioned an expert group to advise on the feasibility of establishing an individual appeals and complaints mechanism for platform content decisions and that group recommended such a mechanism be established.³⁶⁸ The German Federal Court of Justice has also ruled that Facebook must provide every individual user notice if a post is deleted, reasons for that deletion, and an opportunity to reply.³⁶⁹ The Indian government has passed a regulation creating government-appointed “Grievance Appellate Committees” to hear user challenges to company decisions, creating a government-run mechanism resembling the Board.³⁷⁰ Thus, the idealized due process model of governance which the Board epitomizes can be found in many contexts, and so the Board experiment holds important lessons for these other experiments in platform governance. This Part turns to those lessons.

First, I give a realist account of the Board’s incentives to show how institutional incentives have played an underappreciated role in the Board’s work. The starting point is understanding that the Board is a servant of two masters: its creator, Meta, at whose mercy it exercises any power at all, and the public, whose approval it was created to court. The Board’s precarious status depends on walking a tightrope between these two sets of interests. The Board’s decisions are therefore never simply a judgment about the merits of any individual case — they are always, consciously or not, also a political judgment of how best to manage the (often conflicting) interests of these stakeholders. This leads the Board to prioritize bolstering its own reputation over actual impact.

Second, I examine the political environment within which the Board operates to ask why this has incentivized the outcomes that it has. Often, the relief that someone is doing something has overtaken the need to carefully examine exactly what that something is and allowed the Board to reap rewards for pursuing a formalistic style of governance.

Finally, I suggest that what this shows is that legitimacy is an ambiguous indicator of successful governance. Sociological legitimacy is sought after by those in power but is not a standalone good. At best, it

367. TEX. BUS. & COM. CODE § 120.103 (West 2021); FLA. STAT. § 501.2041 (2021).

368. IRISH DEP’T OF TOURISM, CULTURE, ARTS, GAELTACHT, SPORT & MEDIA, REPORT OF THE EXPERT GROUP ON AN INDIVIDUAL COMPLAINTS MECHANISM: MAY 2022, at 2 (Sept. 2022).

369. Press Release, Bundesgerichtshof, Bundesgerichtshof Zu Ansprüchen Gegen Die Anbieterin Eines Sozialen Netzwerks, Die Unter Dem Vorwurf Der “Hassrede” Beiträge Gelöscht Und Konten Gesperrt Hat (July 29, 2021), <https://www.bundesgerichtshof.de/SharedDocs/Pressemitteilungen/DE/2021/2021149.htm> [https://perma.cc/2ZHA-3V33].

370. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2022, G.S.R. 794(E), (India), <https://egazette.nic.in/WriteReadData/2022/239919.pdf> [https://perma.cc/C8CV-YRV7].

is something that can be harnessed to enact substantial reforms. At worst, it can be used to breed complacency about the need for such reforms altogether. Accordingly, we should be far more skeptical about generalized calls for more “legitimate” content moderation governance and instead articulate more specific normative ends.

A. The Importance of Institutional Incentives

The Board’s institutional incentives have played an underappreciated role in dictating the Board’s work and tempering its material impact.³⁷¹ For the Board, as a self-regulatory institution with no legal or other independent authority, and no “sword or . . . purse” of its own,³⁷² its power lies in establishing sociological legitimacy. But the fact that the Board is a creature of self-regulation makes its goal of establishing its legitimacy more challenging, because it was designed by and continues to exist only at the whims of the entity it was created to constrain. Of course, Meta wants the Board to succeed — it did not establish the Board to fail. But it wants the Board to succeed in very particular ways, and this skews the Board’s work.

Meta and the Board’s interests are aligned to an extent. Both want the experiment to work, and both want the Board’s influence to grow.³⁷³ But Meta’s goodwill towards the Board will surely not be infinite. Ultimately, Meta only cares about the Board’s legitimacy and influence to the extent that it redounds to Meta’s own benefit. The goodwill of the regulated entity would not necessarily matter as much in most oversight relationships — but it matters a great deal when the oversight institution exercises power at the grace of the entity being overseen.

It is therefore important to remember Meta’s actual interests in the Board experiment. Meta did not create the Board in order to get yet another form of substantive input on its content moderation decisions, but in order to shift public perception.³⁷⁴ The substantive work of the Board is of little benefit to Meta, which regularly consults with stakeholders to get the perspective of experts and affected communities.³⁷⁵ The distinctive feature of the Board is not the nature of the expertise that the Board provides, but the fact that it provides feedback in a

371. This Section will discuss the Board and its members somewhat as a monolith. Of course, different members will have different values, incentives, and goals. This Section does not engage in individual speculation but is interested in more general and generalizable dynamics.

372. THE FEDERALIST NO. 78 (Alexander Hamilton).

373. See *supra* Section II.B.

374. See Douek, *supra* note 127, at 553–54; Kadri, *supra* note 44, at 189.

375. *How Meta Engages with Stakeholders*, META TRANSPARENCY CTR. (Jan. 18, 2023), <https://transparency.fb.com/policies/improving/stakeholders-engaging-in-policy-development/> [https://perma.cc/69TS-T4VJ].

public-facing way. It is the performance of oversight that is important to Meta. The return Meta wants for its \$280 million investment is enhanced perceived legitimacy of its community standards and reduced costs, both commercial and reputational, of making unpopular decisions. Meta's support of — and obedience to — the Board therefore is contingent on these returns being greater than the costs of complying with the Board. If that compliance becomes too costly, there is no way of forcing Meta to continue to pay attention to the Board.

This makes public approval extremely important to the Board too, not only for personal reputational reasons but also for institutional ones. The fact that Meta could start to ignore the Board tomorrow means that the Board's power and continued existence depend on maintaining a baseline of public support. The Board therefore needs to constantly court public approval in order to increase the reputational costs to Meta of ignoring the Board's recommendations altogether.

But this creates a predicament for the Board because public faith in the Board relies on the Board looking like a meaningful accountability mechanism and not a mere sham. The Board therefore needs to look tough on Meta to garner public approval. But if the Board is too tough on Meta, then Meta will simply ignore it, which would set off a downward spiral, as it would make the Board look weak, which would cause the Board to lose public approval, which makes it easier for Meta to ignore it in future. Thus, to be successful, the Board must walk a fine line between appearing to bring Meta to heel while not getting too far out over its own skis. It must hold Meta accountable, but also know its own limits. As one Board member explained, "one of the dilemmas of the initial stages of the Board is the feeling that . . . if we play too hardball, they may shut us out entirely. And so we have to . . . develop a kind of collaborative *modus operandi*; we can't be too tough, lest this experiment kind of fail on inception."³⁷⁶

Again, the Board is in a position similar to a court in an authoritarian regime. Martin Shapiro explains the legitimacy paradox in the context of courts trying to rein in authoritarian governmental abuses in the following way: if a court challenges a regime to the extent that it gets ignored, the court loses its legitimacy, and if it doesn't challenge the regime at all it also loses its legitimacy.³⁷⁷ But even "[i]f they manage things just right and maintain some perceived legitimacy, they lend that legitimacy to the authoritarian regime of which they are a part precisely because they are a part of it."³⁷⁸ This means that the Board is necessarily co-opted into bolstering Meta's own legitimacy, not merely its

376. NBC News, *PEN America CEO Says Facebook Is Not Always Responsive to Its Oversight Board*, YOUTUBE, at 2:34 (June 27, 2023), <https://www.youtube.com/watch?v=ZxLvXEKs-NY> [https://perma.cc/2ACG-T8A3].

377. Shapiro, *supra* note 90, at 334.

378. *Id.*

own. And it also means that the Board cannot risk the weakness that comes with being too publicly controversial.

Ultimately, this explains the Board's reluctance to meaningfully engage with the project of developing a normative framework for its decisions, focusing instead on easy-to-measure metrics of success rather than necessarily meaningful ones, and its pattern of sidestepping hard calls. The Board is not incentivized to provide answers, so much as it is incentivized to slap Meta on the wrist.

Take the development of a normative framework for thinking about the private regulation of speech in a rights-respecting and coherent way. Such a task is hard and will be controversial. As the Board is confronted with this task, there is little the Board can fall back on to justify its decisions beyond first principles, making it especially vulnerable to criticism for its own normative choices. Instead of showing its work, therefore, the Board simply presents many of its conclusions as compelled by IHRL, when in fact what IHRL would require in the context of most content moderation decisions is almost entirely under-specified and indeterminate. That is, the Board is making political choices but obscuring them. And it is deferring the work of creating a principled framework that creates predictability and can reconcile the various tensions in its work.

This is a form of “[s]trategic avoidance — postponing decision of contentious issues that might threaten a [decision-maker]’s institutional viability”³⁷⁹ — that is often used by newer institutions uncertain about their power. For such institutions, “[a]voidance is a means of cultivating . . . legitimacy. A [decision-maker] can simply avoid deciding contentious, politically divisive issues that, by creating powerful opponents with the capacity to rein in (or oppose) the [decision-maker]’s actions, could threaten its institutional legitimacy.”³⁸⁰ This also gives the decision-maker time to shore up support for its authority and decisions before engaging in confrontation.³⁸¹ Through this lens, the Board’s actions can be seen as a strategy for the Board reigning in the ambitions of its decisions and muting their political significance is a strategic move in order to bide its time and try to establish a base of public support before becoming more confrontational.

This also explains the Board’s emphasis on doing so much public relations and institutional promotion — whether through its own corporate communications team and accounts, or through the regular appearance of members of the Board at various events with stakeholders in digital governance present — even as the Board’s caseload continues to dwindle, and it keeps missing deadlines for its work. Convincing the

379. Delaney, *supra* note 264, at 4.

380. *Id.* at 10.

381. Dixon & Issacharoff, *supra* note 340, at 687.

public of the institution's worth is as important to the Board as the actual work of governing.

The Board's practice of public relations and strategic avoidance has so far seemingly served it well. Whether the Board could have achieved the power building that it has while being less avoidant is of course impossible to know. But as time goes on, such an approach gets riskier. If the Board continues to duck all the most difficult questions and neglect its job of building a more robust normative framework for content moderation decisions, it will consign itself to irrelevance. Meta will no longer refer its most difficult decisions to the Board if it can expect the Board to simply keep dodging the hardest questions. That is, if the Board becomes "known to avoid politically divisive issues, it may lose its authority to decide controversial cases."³⁸² And the Board's practice of continually recommending more due process mechanisms in place of giving substantive answers to questions will be less popular in a less favorable economic environment in which many platforms are cutting back, not further investing, in trust and safety.³⁸³

But perhaps most importantly, the public's interest and trust in the institution should wane if it cannot more convincingly demonstrate its added value. The failure to give compelling and comprehensive reasons for its decisions and build out its normative framework is a failure to live up to its potential. It undermines the predictability of the Board's decisions and their value as a guide for other decision-makers. It also makes it increasingly hard to credit the Board members' expertise and unique qualifications to be making the kinds of decisions with which they are tasked if they do not display that expertise in the decisions themselves. And it perhaps suggests the real limits of self-regulation in general. If the Board never feels emboldened enough to stand up to Meta or public opinion in high-stakes moments, it might be a sad indication of the constraints on what self-regulatory institutions might be able to achieve. Yet, despite all this, public interest and trust in the Board seems to be trending up, not down.

B. The Revealed Preference for Performative Governance

The institutional incentives to perform solemn governance without rustling too many feathers exist because the formalistic approach to governance *is working*. That is one of the most important lessons of the Board experiment — the reception of the Board suggests that the

382. Delaney, *supra* note 264, at 15.

383. See J.J. McCorvey, *Tech Layoffs Shrink 'Trust and Safety' Teams, Raising Fears of Backsliding Efforts to Curb Online Abuse*, NBC NEWS (Feb. 10, 2023), <https://www.nbcnews.com/tech/tech-news/tech-layoffs-hit-trust-safety-teams-raising-fears-backsliding-efforts-rcna69111> [https://perma.cc/5AYS-VE6F].

pursuit of legitimacy through procedural mechanisms can be effective.³⁸⁴ There is something about the court-like model that appeals to stakeholders. It invokes the familiar state-based model of speech governance and no doubt benefits from legitimacy by association. It provides a focal point for people to express their outrage, while also giving the appearance of finality once a decision is issued. It provides the show of Meta being reprimanded (“Oversight Board Criticizes Meta for Preferential Treatment”³⁸⁵ reads one headline; “Facebook Is Rebuked by Oversight Board Over Transparency on Treatment of Prominent Users”³⁸⁶ says another), even if the real-world costs to Meta of such a reprimand are low. And the implementation of recommendations means that *something* is being done to reform the current systems. These all seem to scratch the itch of dissatisfaction with the prior regime of Meta’s unilateral and completely unaccountable system of governance.

It’s important to be clear: the Board is not being performative in the way that its most trenchant critics warned when it was being established. It is not a “sham”³⁸⁷ or mere “spin.”³⁸⁸ It has not convinced lawmakers that legal regulation of Meta is no longer necessary or prevented continued trenchant critique of Meta’s decision-making.³⁸⁹ The Board is meaningfully independent and is apparently prompting Meta to make changes to its systems, often of the kind that civil society and others have been calling for from platforms for years, and indeed that regulators are drafting laws to bring about now. By making recommendations to Meta that bring about some of these reforms through self-regulation, the Board *is* governing, and in exactly the way that many stakeholders have asked for. It is receiving sociological legitimacy precisely because it is far less performative than many cynics predicted. Or, perhaps more specifically, it is performative in a way that is seen as legitimate.

384. See BEN BRADFORD, FLORIAN GRISEL, TRACEY L. MEARES, EMILY OWENS, BARON L. PINEDA, JACOB N. SHAPIRO ET AL., THE JUST. COLLABORATORY, YALE L. SCH., REPORT OF THE FACEBOOK DATA TRANSPARENCY ADVISORY GROUP 33–39 (2019), https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf [<https://perma.cc/6GFS-BC6B>].

385. Adam Satariano & Sheera Frenkel, *Oversight Board Criticizes Meta for Preferential Treatment*, N.Y. TIMES (Dec. 6, 2022), <https://www.nytimes.com/2022/12/06/technology/meta-preferential-treatment.html> [<https://perma.cc/GC4B-Z3C7>].

386. Schechner, *supra* note 216.

387. McNamee & Ressa, *supra* note 33.

388. Olivia Solon, *While Facebook Works to Create an Oversight Board, Industry Experts Formed Their Own*, NBC NEWS (Sept. 25, 2020), <https://www.nbcnews.com/tech/tech-news/facebook-real-oversight-board-n1240958> [<https://perma.cc/RHQ5-SA3U>].

389. For example, Meta did not seem to receive less criticism than Twitter (subsequently known as X) and YouTube for its decision to reinstate Donald Trump’s accounts, even though Meta went through the extensive process of getting the Board to review the decisions while the other two companies made sudden and seemingly arbitrary decisions accompanied by little public reasoning.

C. Sociological Legitimacy as a Poor Marker of Success

Ultimately, the Board should be a reminder that legitimacy is not a standalone good, and social media governance should not be thought to be successful based simply on whether it attains sociological legitimacy amongst elite constituencies. This kind of legitimacy obviously matters to companies like Meta, but it does not translate into material benefits for others.³⁹⁰ Worse, focusing on legitimacy as the metric that needs to be maximized can divert attention for the need for other kinds of reforms.³⁹¹ And, as in authoritarian regimes, a veneer of legitimacy can provide cover for continued engagement with an otherwise objectionable institution.³⁹² What this Article has shown is that institutions can establish legitimacy even as they neglect core parts of their job.

Ultimately, calls for “more legitimate” social media governance are both accurate and too abstract to be useful as a guide for institutional behavior. They respond to the understandable intuition that the prior regime of unilateral power over content moderation is normatively undesirable, but they avoid the much harder and ultimately inherently contestable question of what would be better. The lack of a specific vision of what content moderation governance should be aiming for has created the conditions in which the Board pursues the performance of governing as much, if not more, than its substance.

In many ways this should not be surprising. This dynamic of the pursuit of procedural reforms in the name of “good governance” or rights protection and potentially at the expense of effective governance has been seen in areas as diverse as administrative law³⁹³ and international law.³⁹⁴ But content moderation governance should learn from those mistakes, not replicate them.

The point is not that there are some obviously correct criteria of good governance that the Board should be using instead. The point is that the Board was set up in order for debates about what those criteria should be to take place, and it has instead eschewed those difficult questions or tried to forestall such debate rather than encourage it. The Board has not defined the constituency it is serving or the yardsticks by which it measures its own success. It highlights metrics in its public

390. Making this point in another context, see Ryan D. Doerfler & Samuel Moyn, *Democratizing the Supreme Court*, 109 CAL. L. REV. 1703, 1745 (2021) (“[M]ost approaches to legitimacy define it in terms of partisan neutrality rather than rights protection.”).

391. See SAMUEL MOYN, TESTIMONY BEFORE THE PRESIDENTIAL COMMISSION ON THE SUPREME COURT OF THE UNITED STATES 11 (2021), <https://www.regulations.gov/comment/PCSCOTUS-2021-0001-0118> [<https://perma.cc/EGQ9-EDRT>].

392. Moustafa & Ginsburg, *supra* note 89, at 6.

393. See Bagley, *supra* note 311, at 369.

394. See David Kennedy, *The International Human Rights Movement: Part of the Problem?*, 15 HARV. HUM. RTS. J. 101, 110 (2017) (“The strong attachment of the human rights movement to the legal formalization of rights and the establishment of legal machinery for their implementation makes the achievement of these forms an end in itself.”).

communications — like the number of times the Board “overturns” Meta or how often Meta “takes action” based on the Board’s recommendations — that are poor proxies for actual impact, but good proxies for the kinds of things that attract headlines. It doesn’t encourage debate or disagreement with its decisions, giving the briefest and most abstract possible account of minority views or disagreement within the Board itself in its decisions.

The Board experiment also makes clear that oversight institutions cannot and do not operate in a vacuum, and they respond to the incentives created by their political environment. This means that setting up oversight institutions is only the first step in creating an ecosystem of accountability for content moderation. Watchdogs also need to be watched and made to show their work. Accountability and oversight need to be dynamic and ongoing processes involving a broad base of stakeholders. The Board has not so far been subjected to the level of critical examination commensurate with its status and influence. Part of the problem is surely lack of competition — the Board currently has a monopoly as the only example of public-facing content moderation oversight and precedent-based decision-making. Perhaps the creation of other oversight institutions, such as those that will spring up to serve companies who need to offer recourse to an independent third-party arbiter under Europe’s DSA, will spur improvements. But part of the problem is also that content moderation discourse has always been more interested in spectacle than substance. For years, content moderation debates have focused on anecdotal and individual decisions, rather than interrogating the underlying systems and the pragmatic trade-offs that such speech ecosystems entail.³⁹⁵ This is an environment that will naturally focus on headlines about the Board rapping Meta over the knuckles, rather than pushing for further answers about the normative assumptions and trade-offs that underpin the Board’s decisions, too. The Board could and should be a mechanism for changing that dynamic, rather than playing into it.

V. CONCLUSION

Has the Board improved content moderation governance? We might say, “compared to what?” To say that the Board is better than nothing, and achieved some things, is a start. But it is harder to evaluate whether the Board’s achievements have been meaningful, whether they are commensurate with the level of investment and attention they have been given, or whether they are the best we can hope for from content moderation oversight. Importantly, the answers to those questions

395. Evelyn Douek, *Governing Online Speech: From “Posts-As-Trumps” to Proportionality and Probability*, 121 COLUM. L. REV. 759, 808–09 (2021).

require much more critical engagement with the Board's work than has generally been undertaken so far, more attention to the politics and incentive structures of the Board itself, and greater demands for proof of work. When the Board was being created, the discourse was dominated by concerns about Meta's incentives and how to ensure that the Board was sufficiently insulated from its creator to serve as a meaningful accountability mechanism. But this Article has argued that the Board's own institutional incentives and political environment are impacting its work in significant and underappreciated ways, and it is time to pay more attention to them.

This Article's critiques of the Board should not therefore be understood as critiques of the Board alone. Instead, this is also an indictment of broader content moderation discourse that — after first dismissing the Board entirely — has come to assume the Board's significance and accept its influence rather uncritically, while failing to ask enough questions about what, exactly, the Board is or should be trying to achieve.