

THE META OVERSIGHT BOARD'S HUMAN RIGHTS FUTURE

Laurence R. Helfer[†] & Molly K. Land^{††}

Responses to the new Oversight Board created by Facebook (now Meta) have run the gamut from enthusiastic to overtly suspicious. Many observers are highly skeptical of the Board's ability to hold Meta accountable or to protect the company's users. Much of this skepticism is based on a misplaced analogy between the Board and domestic courts. We argue, in contrast, that the core challenges and opportunities that the Oversight Board faces are similar to those confronted by international human rights tribunals. Analyzing the Oversight Board as a de facto human rights tribunal sheds light on the strengths and weaknesses of its structure, decision-making, and potential future trajectory. Seen from this perspective, key design features that commentators have criticized are, in fact, strengths the Board is already using to expand its authority, develop human rights norms, and influence efforts to regulate social media platforms.

This Article is the first to examine the human rights origins of the Oversight Board, the similarities between the new body and international human rights tribunals, and how the Board is using human rights standards to pressure Meta to improve its content moderation policies and to inform ongoing efforts to regulate social media companies. We complement this analysis with a range of recommendations for the Board to become even more effective, as well as explore

[†] Harry R. Chadwick, Sr. Professor of Law, Duke University. This Article is written in the author's personal capacity. It does not express opinions on behalf of the U.N. Human Rights Committee, on which the author currently serves as a member. This Article analyzes Oversight Board decisions and related developments through May 1, 2023.

^{††} Catherine Roraback Professor of Law and Human Rights, University of Connecticut School of Law. We are grateful for feedback and comments from Agustina del Campo, Stefania Di Stefano, evelyn douek, David Kaye, and Richard Wingfield, as well as participants in the Boston College Law School and University of Minnesota Law School faculty workshops, the European Society of International Law's Interest Groups on International Courts and Tribunals and on Business and Human Rights, and the City University of London conference, "The Facebook/Meta Oversight Board: International and Regional Law Applications." We are also grateful for the opportunity to present our findings to Meta's Human Rights and Governance teams. Excellent research assistance was provided by Noah Buyon, David Del Prete, Ben Greene, Sophia Nelson, and Tatiana Varanko.

potential risks and challenges, including backlash, whitewashing, and negative spillover. Although a sanguine vision of the Board's future is by no means assured, when viewed in light of the experience of international human rights tribunals, we conclude that the Board has the potential to serve as a meaningful check on Meta and to significantly advance the promotion and protection of human rights online.

TABLE OF CONTENTS

INTRODUCTION	2235
I. THE OVERSIGHT BOARD AS A HUMAN RIGHTS TRIBUNAL.....	2241
A. <i>The Influence of Human Rights on the OB's Origins and Design Features</i>	2241
1. Origins.....	2241
2. Jurisdiction and Standing	2245
3. Applicable Law and References to Human Rights.....	2246
B. <i>Shortcomings in Design</i>	2247
C. <i>A Better Analogy: International Human Rights Tribunals</i>	2250
1. Similarities	2250
2. Differences	2254
D. <i>The Impact of International Human Rights Tribunals: Accountability Promotion and Norm Development</i>	2256
1. Promoting Accountability.....	2258
2. Developing International Norms	2259
II. THE OVERSIGHT BOARD'S HUMAN RIGHTS FUTURE	2261
A. <i>Accountability Promotion</i>	2261
1. Jurisdiction	2261
2. Standards of Review	2266
3. Applicable Law.....	2269
4. Evidentiary Issues	2272
5. Compliance.....	2273
B. <i>Norm Development</i>	2275
1. The Oversight Board as Norm Generator.....	2276
2. Norm Development in OB Caselaw.....	2277
3. The Impact of Norm Generation on Human Rights.....	2281
a. Norms as Resources for Mobilization.....	2282
b. Dialogue Promotion.....	2283
c. Norm Diffusion.....	2285
III. RISKS AND CHALLENGES	2288
A. <i>Pushback and Backlash</i>	2289
B. <i>Whitewashing</i>	2291
C. <i>Indeterminacy and Legitimacy</i>	2294

2023]

META OVERSIGHT BOARD

2235

D. <i>Negative Spillover</i>	2297
CONCLUSION.....	2299

INTRODUCTION

In late 2020, Facebook (now Meta)¹ created an Oversight Board (“OB” or “Board”) to review challenges to the social media platform’s decisions to remove user posts and to advise the company on improving its content moderation policies. At the time, the Board drew widespread notice and scrutiny from civil society groups and commentators, whose responses have run the gamut from enthusiastic to overtly suspicious. The Oversight Board has recently received fresh attention as governments adopt new initiatives to regulate internet platforms.² Several of these initiatives impose liability on social media companies for retaining or removing user content and establish procedures for users to contest whether such content is illegal or violates the companies’ terms of service.³

Does the Meta Oversight Board represent the first wave of social media companies becoming more accountable to their users and other stakeholders? Does the Board provide a model for government regulators and other internet platforms to move toward greater transparency in content moderation? Or is it merely a fig leaf that seeks to ameliorate

¹ In 2021, Facebook officially changed its name to Meta. *Introducing Meta: A Social Technology Company*, META (Oct. 28, 2021), <https://about.fb.com/news/2021/10/facebook-company-is-now-meta> [<https://perma.cc/JR26-3LQR>]. The name “Facebook” continues to be the name for the platform and the app, however. For the sake of continuity, this Article uses “Meta” to refer to the company both before and after the name change.

² See, e.g., Molly K. Land & Laurence Helfer, *Value Pluralism and Human Rights in Content Moderation*, LAWFARE (Oct. 27, 2022, 8:16 AM), <https://www.lawfareblog.com/value-pluralism-and-human-rights-content-moderation> [<https://perma.cc/W3Q2-J4TF>] (describing potential conflicts between new European Union regulations and U.S. law, including the First Amendment, and suggesting ways to mitigate them); David Morar & Bruna Martins dos Santos, *The Push for Content Moderation Legislation Around the World*, BROOKINGS: TECHTANK (Sept. 21, 2020), <https://www.brookings.edu/blog/techtank/2020/09/21/the-push-for-content-moderation-legislation-around-the-world> [<https://perma.cc/FJN3-6F3M>] (describing regulatory initiatives in Brazil, France, Germany, and the United States).

³ The most important of these initiatives is Regulation 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 [hereinafter DSA]. The DSA establishes a three-tiered system for contesting a decision made by an online platform as to whether user content is illegal or violates the platform’s terms and conditions: internal complaints, out-of-court dispute settlement, and redress in court. For a comparison of these procedures with the Oversight Board, see David Wong & Luciano Floridi, *Meta’s Oversight Board: A Review and Critical Assessment*, MINDS & MACHS. (Oct. 24, 2022), <https://doi.org/10.1007/s11023-022-09613-x> [<https://perma.cc/A4F8-QAU6>].

Meta's many public relations challenges while obscuring its human rights harms?

Attempts to answer these questions are hampered by a lack of agreement about how to conceive of this new body. Not quite a corporate entity and not really a court, the Oversight Board is a body of nearly two dozen experts drawn from geographically, culturally, and professionally diverse backgrounds, funded by an independent trust created under Delaware law, empowered by Meta to make decisions on content that affect billions of individuals worldwide. First proposed as a “quasi-judicial unit with autonomous powers” inspired by constitutional law,⁴ the idea of an independent review mechanism captured the imagination of CEO Mark Zuckerberg, who floated the idea of creating a “Supreme Court of Facebook”⁵ to help the company make decisions about speech on its platforms.⁶ Since its launch, commentators have compared the Oversight Board to U.S. courts,⁷ administrative tribunals,⁸ constitutional

⁴ Steven Levy, *Why Mark Zuckerberg's Oversight Board May Kill His Political Ad Policy*, WIRED (Jan. 28, 2020, 2:00 PM), <https://www.wired.com/story/facebook-oversight-board-bylaws> [<https://perma.cc/SR9H-47YS>]; Mark Sullivan, *Exclusive: The Harvard Professor Behind Facebook's Oversight Board Defends Its Role*, FAST CO. (July 8, 2019), <https://www.fastcompany.com/90373102/exclusive-the-harvard-professor-behind-facebooks-oversight-board-defends-its-role> (last visited May 20, 2023).

⁵ Levy, *supra* note 4.

⁶ Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK, <https://www.facebook.com/notes/751449002072082> (May 5, 2021) (last visited May 20, 2023); 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/AT3Q-2D8B>].

⁷ Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2476–77 (2020); Jeremy Lewin, *Facebook's Long-Awaited Content 'Supreme Court' Has Arrived. It's a Clever Sham*, GUARDIAN (Mar. 17, 2021, 6:23 AM), <https://www.theguardian.com/commentisfree/2021/mar/17/facebook-content-supreme-court-network> [<https://perma.cc/GUQ5-66XS>].

⁸ Klonick, *supra* note 7, at 2476–77; We the People, *Trump and the Facebook Oversight Board*, NAT'L CONST. CTR., at 03:32 (May 6, 2021), <https://constitutioncenter.org/interactive-constitution/podcast/trump-and-the-facebook-oversight-board> [<https://perma.cc/W4ZV-WSRA>].

courts,⁹ private arbitral bodies,¹⁰ grievance mechanisms,¹¹ and self-regulatory initiatives.¹²

Each of these analogies, however, fails to account for the Board's distinctive design features, the circumstances of its creation, and its potential for developing international norms. Analogies to domestic courts, for example, inaccurately assume that the OB operates within a preexisting and stable legal system with established checks and balances and enforcement powers. Analogies to contractual or self-regulatory mechanisms overlook the public law aspects of the Board's work or the range of stakeholders that are affected by its decisions.

Commentators also disagree over what this new body should do. Some argue that the Board's primary goal should be to increase Meta's accountability to users¹³ or to provide a mechanism for individual redress.¹⁴ Others have emphasized the OB's role in curbing the company's

⁹ evelyn douek, *Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility*, 21 N.C. J.L. & TECH. 1, 3 (2019); 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/JA7K-PWZ7>]; Oreste Pollicino & Giovanni De Gregorio, *Shedding Light on the Darkness of Content Moderation*, VERFASSUNGSBLOG (Feb. 5, 2021), <https://verfassungsblog.de/fob-constitutionalism> [<https://perma.cc/Y5LD-WYXE>]. For a discussion of the use and impact of the Supreme Court metaphor for the Board, see Josh Cowls, Philipp Darius, Dominiquo Santistevan & Moritz Schramm, *Constitutional Metaphors: Facebook's 'Supreme Court' and the Legitimation of Platform Governance*, NEW MEDIA & SOC'Y (Apr. 5, 2022), <https://doi.org/10.1177/14614448221085559> (last visited May 20, 2023).

¹⁰ See Klonick, *supra* note 7, at 2477; Daniel Milo, *Is Facebook's Oversight Board a Much-Needed Fix for the Platform's Many Ills or a PR Stunt?*, NEW EUR. (Feb. 25, 2021, 3:39 PM), <https://www.neweurope.eu/article/is-facebooks-oversight-board-a-much-needed-fix-for-the-platforms-many-ills-or-a-pr-stunt> (last visited May 20, 2023); Imre Szalai, *Facebook's Oversight Board, Trump, and Arbitration*, ARBITRATIONUSA (May 5, 2021), <https://arbitrationusa.com/2021/05/facebooks-oversight-board-trump-and-arbitration> [<https://perma.cc/2LFD-CMWU>].

¹¹ Blayne Haggart & Clara Iglesias Keller, *Democratic Legitimacy in Global Platform Governance*, 45 TELECOMMS. POL'Y, July 2021, at 1, 8–9; Jenny Domino, *The Facebook Oversight Board as Operational-Level Grievance Mechanism*, CAMBRIDGE UNIV. PRESS: CAMBRIDGE CORE BLOG (Apr. 1, 2020), <https://www.cambridge.org/core/blog/2020/04/01/the-facebook-oversight-board-as-operational-level-grievance-mechanism> [<https://perma.cc/W2VD-QUE7>]; see Ruby O'Kane, *Meta's Private Speech Governance and the Role of the Oversight Board: Lessons from the Board's First Decisions*, 25 STAN. TECH. L. REV. 167, 200–01 (2022).

¹² Chinmayi Arun, *Facebook's Faces*, 135 HARV. L. REV. F. 236, 244 (2022); see also Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT'L L.J. 41, 90–93 (2020); Levy, *supra* note 4.

¹³ Klonick, *supra* note 7, at 2478.

¹⁴ Gabriella Casanova, Carlos Lopez & Sam Zarifi, *Some Questions Regarding Facebook's Oversight Board and Remediation of Human Rights Impacts (Part I)*, OPINIOJURIS (Mar. 3, 2020), <http://opiniojuris.org/2020/03/03/some-questions-regarding-facebooks-oversight-board-and-remediation-of-human-rights-impacts-part-i> [<https://perma.cc/H8SR-QH3X>]; Rishi Gulati,

adverse human rights impacts or as a forum for negotiating the rules governing online speech.¹⁵ The Charter establishing the Oversight Board describes its purpose as protecting free expression “by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook’s content policies.”¹⁶ Yet the OB also calls itself an “independent grievance mechanism”¹⁷ that “promote[s] the rights and interests of users.”¹⁸

The Board’s design raises significant doubts about its ability to accomplish any of these objectives. The OB’s authority is limited: It can review Meta’s decisions to remove or allow individual pieces of content on Facebook and Instagram that are contested by users or submitted for review by the company itself. Additionally, it can evaluate Meta’s content moderation policies, but its recommendations are only advisory. The OB is also hampered by jurisdictional and functional constraints. Meta’s news feed ranking and political advertising are outside the Board’s remit, as are the company’s decisions to comply with local law.¹⁹ In addition, the Board has been issuing between twelve and sixteen decisions each year. Given the volume of content shared on the platform, such a limited number of individual cases seems unlikely to have much impact.²⁰ Lastly, the Board is perpetual in theory but time limited in practice, since its existence likely depends on continued funding from Meta.

These limitations are daunting. They are not, however, all that new. The central challenges that the Oversight Board faces—as a body of limited, consent-based jurisdiction and constrained authority, facing overwhelming demand, whose existence, legitimacy, and efficacy depend in large part on the continued support of its creator—are similar to the challenges confronted by international human rights courts and quasi-judicial review bodies such as U.N. human rights treaty bodies, the European Court of Human Rights, and the Inter-American Commission

Meta’s Oversight Board and Transnational Hybrid Adjudication—What Consequences for International Law? 20 (Berlin Potsdam Rsch. Grp., KFG Working Paper No. 53, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4048587 [<https://perma.cc/G2MF-J9GX>].

¹⁵ douek, *supra* note 9, at 7; Haggart & Keller, *supra* note 11, at 9.

¹⁶ OVERSIGHT BD., OVERSIGHT BOARD CHARTER intro. (2019) [hereinafter OVERSIGHT BOARD CHARTER], <https://www.oversightboard.com/governance> [<https://perma.cc/82J7-VULC>].

¹⁷ Former President Trump’s Suspension, No. 2021-001-FB, 2021 WL 4822219, at *8 (May 5, 2021); South Africa Slurs, No. 2021-011-FB, 2021 WL 4822229, at *3 (Sept. 28, 2021). For all Oversight Board case decisions discussed in this Article, see *Decision*, OVERSIGHT BD., <https://www.oversightboard.com/decision> [<https://perma.cc/73KP-6N96>].

¹⁸ OVERSIGHT BD., OVERSIGHT BOARD ANNUAL REPORT 2021, at 3 (2021) [hereinafter OVERSIGHT BOARD ANNUAL REPORT], <https://oversightboard.com/attachment/425761232707664> [<https://perma.cc/KB7J-LNFS>].

¹⁹ douek, *supra* note 9, at 38, 45. *But see infra* text accompanying notes 142, 148–57.

²⁰ evelyn douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 569 (2022).

on Human Rights (“international human rights tribunals” or “tribunals”).²¹ We argue that, like international human rights tribunals, the Board can contribute most effectively to the international human rights system not by correcting errors or providing remedies in individual cases but by pressuring Meta to improve its processes and by developing the law.

This Article sketches a comprehensive vision for the Oversight Board as a de facto international human rights tribunal. It does so in three steps. We first explore the influence of human rights law and tribunals on the Board’s origins, its design features, and operations. Previous accounts of the OB have overlooked how these sources affected the process leading to the Board’s creation, the choices regarding the OB’s review powers, and the legal norms it applies.

Second, we consider the structural and functional analogies between the OB and international human rights tribunals. We draw on the extensive experience of these tribunals over the last several decades to highlight the many similarities between the Oversight Board and the tribunals, while also noting salient differences.²² Comparing the OB to these bodies sheds light on the strengths and weaknesses of the Board’s structure, its decision-making processes, and its potential future trajectory.²³ The similarities between the two sets of institutions reveal that key OB design features that some observers have criticized—such as its ability to issue only narrow binding rulings while designating its policy guidance as nonbinding—are actually strengths that the Board can use to incrementally expand its authority over time. We also review OB decisions that suggest that the Board is beginning to emulate the course previously charted by international human rights tribunals, for example by issuing detailed policy recommendations and pressing Meta to provide

²¹ We use the phrase “international human rights tribunals” to include all judicial and quasi-judicial bodies established by international agreements or international organizations to evaluate the human rights practices of states. These bodies operate within the United Nations and at the regional and subregional levels. The comparison between the Board and these bodies as a whole necessarily glosses over the differences between the tribunals themselves, including differences in the challenges they have faced and the strategies they have developed to overcome those challenges.

²² International human rights tribunals are not the only possible analogy. Another may be the World Bank Inspection Panels, independent recourse mechanisms created in 1993 to enable private actors to challenge the World Bank’s actions. Panels may only determine whether the Bank’s actions comply with its Operational Policies and Procedures. However, when doing so they can consider human rights norms to the extent those norms are implicitly embedded in the Bank’s policies and procedures. For further discussion, see INT’L BANK FOR RECONSTR. & DEV., ACCOUNTABILITY AT THE WORLD BANK: THE INSPECTION PANEL AT 15 YEARS 74 (2009).

²³ See Anthea Roberts, *Clash of Paradigms: Actors and Analogies Shaping the Investment Treaty System*, 107 AM. J. INT’L L. 45, 56 (2013) (“Framing plays a crucial role in how people conceptualize issues and analyze problems, but its effects are typically covert, as people rarely recognize the role that framing plays in shaping their reasoning.”).

information about its algorithms and communications with governments.

Third, we sketch a vision for the Oversight Board's human rights future. We argue that the OB should develop methodologies similar to those that the tribunals have used to promote accountability and to contribute to human rights norm development. We contend that the 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. We show that the Board is already making significant contributions in this regard—some of which are being cited favorably by other human rights bodies—and we highlight opportunities for the OB to influence the regulation of internet platforms by national governments and the European Union.

This sanguine vision of the Board's human rights future is by no means assured, however. Efforts by the OB to enforce its existing mandate or expand its authority may engender resistance by Meta. The company may also engage in whitewashing, touting its commitment to the Board and to human rights generally in an attempt to avoid more stringent regulation. In addition, skeptics have also questioned the legitimacy of a private monitoring body developing public norms for content moderation that are precise enough to impose meaningful constraints on social media companies while also taking account of the different ways that the value and harm of speech are balanced around the world. There is also a risk of negative spillover if states and companies misuse the Board's normative output in ways that weaken existing international human rights standards. We take these concerns seriously and suggest a number of ways to mitigate them.

This Article contributes to robust ongoing debates about social media content regulation in several ways. First, we analogize the Oversight Board to international human rights tribunals to shed light on the opportunities, risks, and challenges that the OB is facing and is likely to face. Second, we provide the first comprehensive analysis of the Board's contributions to human rights doctrine, including its clarification of how freedom of expression and other human rights apply to private internet platforms in a wide array of contexts. Third, we offer recommendations for the Board to shape its future trajectory as a *de facto* international human rights tribunal, guidance that we hope will also influence other initiatives to address social media harms.

The Article proceeds in three Parts. Part I describes the role of international human rights law and institutions in the Oversight Board's origins and design and identifies key similarities between the Board and international human rights tribunals. Drawing on the experiences of

2023]

META OVERSIGHT BOARD

2241

human rights tribunals, Part II describes two ways in which the Board will have an impact on both Meta and human rights law and explores how these approaches have already emerged in the Board's decisions. Part III discusses the risks and challenges the OB may face in pursuing this trajectory. A brief conclusion summarizes the Article's contributions.

I. THE OVERSIGHT BOARD AS A HUMAN RIGHTS TRIBUNAL

This Part reviews the origins and design features of the Oversight Board and compares and contrasts these features with those of international human rights tribunals.²⁴ This comparison reveals that what many commentators view as weaknesses—in particular, the OB's circumscribed jurisdiction and authority—may actually help the Board promote human rights.

A. *The Influence of Human Rights on the OB's Origins and Design Features*

In this Section, we provide basic information about the Board's origins and functions, explore how international human rights law and tribunals were discussed during the lead up to the Board's creation, and analyze the references to human rights in the Board's founding documents.

1. Origins

The Oversight Board's trajectory from initial conception to functioning institution was a remarkably short two and a half years. The first public statements about the Board occurred in early 2018, including Mark Zuckerberg's reference to a "Supreme Court" that would "ultimately make the final judgment call on what should be acceptable speech."²⁵ The company announced plans for what eventually would be called the Oversight Board in late 2018; the Board's Charter was revised and finalized during 2019; Meta released the OB's Bylaws in January 2020;

²⁴ The design and structure of the Board has been explored thoroughly by Kate Klonick and other scholars, and we do not seek to replicate that work. See Klonick, *supra* note 7.

²⁵ 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/VR2F-M43Q>].

and the inaugural members were appointed by May 2020.²⁶ The Board began operating in the fall of 2020.²⁷

The Board itself does not have a formal legal personality. The Trust Agreement established by Meta creates an irrevocable noncharitable trust under Delaware law.²⁸ The trust is empowered to create a limited liability company that, in turn, has the authority “to establish, administer, and attend to the ongoing operation of the group of individuals who make up the Board Members.”²⁹

The current composition of the Board is diverse along multiple dimensions. Twenty-two members hail from nineteen countries and bring expertise and professional backgrounds in law, human rights, technology regulation, international politics, content moderation, journalism, health communication, intellectual property, management, policy design, and digital rights, among others.³⁰ The Charter provides that “a committee of the board will select candidates to serve as board members based on a review of the candidates’ qualifications and a screen for disqualifications,” and that both Meta and the public may propose candidates.³¹ The staff that supports the Board’s work also brings a broad range of expertise. The Director of Oversight Board Administration, for

²⁶ *Facebook Oversight Board General Documents*, LAWFARE, <https://www.lawfareblog.com/facebook-oversight-board-general-documents> [<https://perma.cc/X3K9-XSVU>]. Meta selected the initial co-chairs of the Board, and the co-chairs and Meta jointly selected the remaining members. Brent Harris, *Oversight Board Membership Selection Process*, META (Sept. 17, 2019), <https://about.fb.com/news/2019/09/oversight-board-membership> [<https://perma.cc/SF4L-RRB5>]. As of April 2023, Meta has withdrawn from the selection of new Board members. *Facebook Parent Meta to Withdraw from Selection of Oversight Board Members*, ECON. TIMES, <https://economictimes.indiatimes.com/tech/technology/facebook-parent-meta-to-withdraw-from-selection-of-oversight-board-members/articleshow/99276442.cms?from=mdr> [<https://perma.cc/VR9F-W7UB>] (Apr. 5, 2023, 10:35 PM).

²⁷ *Facebook Oversight Board General Documents*, *supra* note 26.

²⁸ OVERSIGHT BD., OVERSIGHT BOARD TRUST §§ 1.3–1.4 (2019) [hereinafter OVERSIGHT BOARD TRUST], <https://about.fb.com/wp-content/uploads/2019/12/Trust-Agreement.pdf> [<https://perma.cc/XSR3-79X3>]; see also Lorenzo Gradoni, *Chasing Global Legal Particles: Some Guesswork About the Nature of Meta’s Oversight Board*, EJIL: TALK! (Dec. 30, 2021), <https://www.ejiltalk.org/chasing-global-legal-particles-some-guesswork-about-the-nature-of-metas-oversight-board> [<https://perma.cc/N77M-MNZM>].

²⁹ OVERSIGHT BOARD TRUST, *supra* note 28, § 2.1.

³⁰ As of May 1, 2023, the countries of affiliation are Australia, Brazil, Cameroon, Colombia, Denmark, Egypt, France, Ghana, Hungary, India, Indonesia, Israel, Mexico, Pakistan, Senegal, Taiwan, the United Kingdom, the United States, and Yemen. *Meet the Board*, OVERSIGHT BD., <https://www.oversightboard.com/meet-the-board> [<https://perma.cc/MF2F-BKRQ>].

³¹ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 8. Under the Charter, the Board cannot have fewer than eleven members and “is likely to be forty members.” *Id.* art. 1, § 1. Terms are for three years, and members can serve for a maximum of three terms. *Id.* art. 1, § 3.

example, is the former Executive Director of ARTICLE 19—one of the preeminent civil society organizations on the freedom of expression.³²

The Board’s structure was developed in conjunction with a consultative process that took place between the release of the initial draft Charter in January 2019 and the publication of the final Charter in September of that year. The January draft posed questions as “a starting point for discussion on how the board should be designed and formed.”³³ From late 2018 through mid-2019, Meta held a series of in-person consultative meetings around the world. The company then published a document summarizing the input from workshops, roundtables, and an online survey.³⁴

During this time, civil society groups and scholars increasingly called for social media platforms to look to international human rights law to guide their content moderation policies.³⁵ These demands featured prominently during the OB consultative process. Advocates argued that international human rights law would “provide the oversight board with a set of tools and a ‘common vocabulary’ for addressing and resolving

³² Staff: Thomas Hughes, OVERSIGHT BD., <https://www.oversightboard.com/meet-the-board/thomas-hughes> [<https://perma.cc/V962-U7UA>].

³³ GLOBAL FEEDBACK AND INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS app. at 2 (2019) [hereinafter CONSULTATION APPENDIX], <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf> [<https://perma.cc/MR9M-NJHY>].

³⁴ FACEBOOK, GLOBAL FEEDBACK & INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS 12 (2019) [hereinafter CONSULTATION OUTCOME DOCUMENT], <https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-2.pdf> [<https://perma.cc/G5Z5-3YMK>] (noting consultations with “experts from multiple disciplines in both the private and public sectors” as well as “[s]takeholders from across the geographic, ideological, and policy spectrum—including vocal critics of Facebook”).

³⁵ See David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶¶ 41–43, U.N. Doc. A/HRC/38/35 (Apr. 6, 2018) [hereinafter *Rep. of the Special Rapporteur*]. See generally Evelyn Mary Aswad, *To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?*, 77 WASH. & LEE L. REV. 609 (2020) [hereinafter Aswad, *To Protect Freedom of Expression*]; 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 (2020); Michael Lwin, *Applying International Human Rights Law for Use by Facebook*, 38 YALE J. ON REGUL. BULL. 53 (2020); 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 (2020); Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26 (2018) [hereinafter Aswad, *The Future of Freedom of Expression Online*]; Lorna McGregor, Daragh Murray & Vivian Ng, *International Human Rights Law as a Framework for Algorithmic Accountability*, 68 INT’L & COMPAR. L.Q. 309 (2019).

hard questions around the moderation of online content.”³⁶ They also noted that Meta had already committed to evaluating the human rights impact of some of its policies as a member of the Global Network Initiative.³⁷ Other contributors, however, worried human rights “would not be enough to adjudicate hard cases” and questioned whether they would provide enough clarity to make difficult tradeoffs.³⁸

Two reports commissioned by Meta shed light on the influence of international human rights law and institutions during the consultative process. The first was produced by BSR, a leading business and human rights advocacy organization. Meta asked BSR to review the consistency of “the final charter, bylaws, and operations” of the Board with “human rights-based approaches.”³⁹ The report strongly recommended that Meta “incorporate a firm commitment to international human rights standards into the Oversight Board’s governance charter and bylaws” and “explain the link between human rights and Facebook’s values.”⁴⁰

The second report, co-authored by a constitutional law scholar and Meta’s Head of Product Policy, discussed potential models for the Board’s design.⁴¹ The templates included the Court of Justice of the European

³⁶ Letter from David Kaye, Special Rapporteur on the Promotion and Prot. of the Right to Freedom of Op. & Expression, to Mark Zuckerberg 2–3 (May 1, 2019), https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/Legislation/OL_OTH_01_05_19.pdf [<https://perma.cc/V3CX-7BQE>]; see also Stefania Di Stefano, *The Facebook Oversight Board and the UN Guiding Principles on Business and Human Rights: A Missed Opportunity for Alignment?*, in HUMAN RIGHTS RESPONSIBILITIES IN THE DIGITAL AGE: STATES, COMPANIES AND INDIVIDUALS 93, 111–12 (Jonathan Andrew & Frédéric Bernard eds., 2021).

³⁷ CONSULTATION OUTCOME DOCUMENT, *supra* note 34, at 34. Meta had also referenced international human rights law in its public statements prior to the creation of the Board. See, e.g., Richard Allan, *Hard Questions: Where Do We Draw the Line on Free Expression?*, META (Aug. 9, 2018), <https://about.fb.com/news/2018/08/hard-questions-free-expression> [<https://perma.cc/H8F7-JLYH>].

³⁸ CONSULTATION OUTCOME DOCUMENT, *supra* note 34, at 34–35; evelyn douek & Kate Klonick, *Facebook Releases an Update on Its Oversight Board: Many Questions, Few Answers*, LAWFARE (June 27, 2019, 3:41 PM), <https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers> [<https://perma.cc/Y3BT-BU2Y>]; see also douek, *supra* note 9, at 59; Brenda Dvoskin, *International Human Rights Law Is Not Enough to Fix Content Moderation’s Legitimacy Crisis*, MEDIUM (Sept. 16, 2020), <https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd> [<https://perma.cc/8P7W-DWSA>].

³⁹ BSR, HUMAN RIGHTS REVIEW: FACEBOOK OVERSIGHT BOARD 3 (2019), https://www.bsr.org/reports/BSR_Facebook_Oversight_Board.pdf [<https://perma.cc/A38X-HLW3>]. The BSR report argued that a main function of the Board should be to provide individual users and other rightsholders with remedies. *Id.* at 29.

⁴⁰ *Id.* at 6.

⁴¹ CONSULTATION OUTCOME DOCUMENT, *supra* note 34, at 15–16. The report discussed:

[C]ommunity and institutional mechanisms for decision oversight (including the Swedish Parliamentary Ombudsman, the general features of Civilian Review Boards for

Union, the Inter-American Commission and Court of Human Rights, and the European Court of Human Rights (ECtHR).⁴² Especially noteworthy was a reference to the ECtHR's use of "precedent as a legitimizing tool to substitute for its lack of coercive power in inducing domestic courts to apply its rulings."⁴³ In addition to these two reports, advocacy groups submitted briefing papers that discussed the relevance of international human rights tribunals.⁴⁴

The final version of the OB that emerged from the consultative process explicitly embraces human rights norms and objectives in several ways. Before describing those references, we first outline the OB's jurisdiction and standing rules to provide context for our analysis.

2. Jurisdiction and Standing

Under the Charter, a case is initiated when an individual Meta user (either the person who posted the content or another user who flagged it for review) or the company asks the Board to review a decision to retain or remove content.⁴⁵ Once in receipt of such a request, the Board has jurisdiction to receive submissions from the user, Meta, and third parties; apply Meta's Community Standards and Values;⁴⁶ seek additional

U.S. policing and U.S. corporate board audit committees, hospital and press oversight entities, the Indian Panchayat systems, and the Pacific Coast ILWU labor grievance arbitration process) as well as a range of judicial and quasi-judicial models such as the U.S. court system, the French Court of Cassation, the European Court of Justice, the Inter-American Commission and Court of Human Rights, the European Court of Human Rights, the World Trade Organization Appellate Body, and the Catholic Apostolic Signatura.

CONSULTATION APPENDIX, *supra* note 33, at 138, 140.

⁴² Human rights tribunals were discussed most extensively in connection with the appointment of judges and workload management, as well as the balance of authority between decision-makers and staff. CONSULTATION APPENDIX, *supra* note 33, at 157–58, 162–67.

⁴³ *Id.* at 158.

⁴⁴ For example, a paper submitted by the Center for Studies on Freedom of Expression and Access to Information (CELE), a research center at the Universidad de Palermo in Buenos Aires, Argentina, discussed the U.N. Human Rights Committee and regional human rights courts, as well as the International Court of Justice and bilateral investment tribunals as models for the Board. CONSULTATION APPENDIX, *supra* note 33, at 116, 120.

⁴⁵ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 2, § 1.

⁴⁶ Facebook's Community Standards are a set of policy rationales and rules on specific categories of content—violence and criminal behavior, safety, objectionable content, integrity and authenticity, and respecting intellectual property—that are prohibited or require safeguards to be posted. *Facebook Community Standards*, META, <https://transparency.fb.com/policies/community-standards> [<https://perma.cc/ZN8G-CCWX>]. The company's Values are outlined in the introduction to the Community Standards, where the Value of "voice" is described as "paramount." Meta may limit voice in the service of four other Values: authenticity, safety, privacy, and dignity. *Id.*

information from the company; and provide a written explanation of its decision. In terms of remedies, the OB can direct Meta to retain or remove the content,⁴⁷ or (following an October 2022 amendment to the Bylaws) to retain it with a warning screen.⁴⁸ The Board can also provide non-binding policy guidance “specific to a case decision or upon [the company’s] request, on [Meta’s] content policies.”⁴⁹

The Bylaws narrow the Board’s jurisdiction described in the Charter. Article 2 of the Bylaws excludes five broad categories of content from OB review (“content posted through marketplace, fundraisers, Facebook dating, messages, and spam”) as well as Meta’s decisions involving intellectual property violations and services on several of Meta’s platforms (“WhatsApp, Messenger, Instagram Direct, and Oculus”).⁵⁰ The Bylaws also exclude review of content removed by Meta “following the receipt of a valid report of illegality” or content that is unlawful or criminally unlawful in a connected jurisdiction—issues that we discuss in greater detail below.⁵¹

The Bylaws also create standing requirements. To request review by the Board, a user “must have an active Facebook or Instagram account.”⁵² Initially, the OB could only review “content that has been removed for violations of content policies.”⁵³ In April 2021, Meta authorized the Board to review user challenges to content that the company decided to retain on its platform.⁵⁴

3. Applicable Law and References to Human Rights

The Oversight Board’s founding documents contain multiple references to human rights. Both the Charter and the Delaware Trust that established the limited liability company that created the OB refer to the Board’s purpose as “protect[ing] free expression by making principled,

⁴⁷ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 4.

⁴⁸ OVERSIGHT BD., OVERSIGHT BOARD BYLAWS art. 2, § 2.3.1 (2022) [hereinafter OVERSIGHT BOARD BYLAWS], <https://www.oversightboard.com/sr/governance/bylaws> [<https://perma.cc/AL2C-VRJ3>].

⁴⁹ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 4.

⁵⁰ OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 1.2.1. The Charter and Bylaws explicitly envision the eventual expansion of the Board’s jurisdiction. The Board is actively engaged in discussions with Meta about expanding its jurisdiction to include “user appeals against the company’s decisions in areas such as groups and accounts, as opposed to just individual posts.” OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 67.

⁵¹ OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 1.2.2; *see also infra* Section II.A.1.

⁵² OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 3, § 1.1.

⁵³ *Id.* art. 3, § 1.1.1.

⁵⁴ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 5.

independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook’s content policies.”⁵⁵

Human rights norms are also central to the law that the Board applies. Although the initial draft of the Charter referred only to Meta’s Community Standards and Values,⁵⁶ the final version adds “human rights norms protecting free expression” as a third source of law.⁵⁷ In addition, the Values explain that Meta will “look to international human rights standards” in deciding whether content that otherwise contradicts the company’s Community Standards should nonetheless remain on the platform.⁵⁸

As we describe below, the Board has fully embraced international human rights law in its decisions, recommendations, and other documents. In its first Annual Report, published in June 2021, the Board emphasized its “conviction that Meta will make content moderation decisions in a fairer, more principled way if it bases those decisions on the international human rights standards to which it has committed itself.”⁵⁹

B. *Shortcomings in Design*

Since the Oversight Board’s inception, commentators have identified a range of shortcomings in the design of this new private monitoring body. This Section catalogues these critiques to set the stage for our more favorable assessment that analogizes the OB to an international human rights tribunal.

Perhaps the most trenchant criticism relates to the Board’s highly circumscribed jurisdiction and access rules in individual content moderation cases.⁶⁰ The Board can only review appeals submitted by individual users and offer recommendations specific to those cases or in response to Meta’s requests for policy guidance.⁶¹ There may thus be decisions or policies that the Board cannot consider because no single user has a sufficient interest to challenge them, even though they have

⁵⁵ OVERSIGHT BOARD TRUST, *supra* note 28, § 2.1.

⁵⁶ CONSULTATION APPENDIX, *supra* note 33, at 3.

⁵⁷ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 2, § 2.

⁵⁸ Monika Bickert, *Updating the Values that Inform Our Community Standards*, META (Sept. 12, 2019), <https://about.fb.com/news/2019/09/updated-the-values-that-inform-our-community-standards> [<https://perma.cc/6HJ2-J5ZE>]; Parmar, *supra* note 9.

⁵⁹ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 9.

⁶⁰ *See, e.g.*, evelyn douek, *What Kind of Oversight Board Have You Given Us?*, U. CHI. L. REV. ONLINE (May 11, 2020), <https://lawreviewblog.uchicago.edu/2020/05/11/fb-oversight-board-edouek> [<https://perma.cc/C7R6-S2U7>]; Zahra Takhshid, *Regulating Social Media in the Global South*, 24 VAND. J. ENT. & TECH. L. 1, 32 (2021).

⁶¹ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 4.

significant consequences when viewed cumulatively. In addition, non-Facebook and non-Instagram users cannot submit cases to the Board⁶² despite the fact that content on the platforms may significantly affect them.⁶³ These limitations may inhibit the Board from considering the systemic or cumulative effects of Meta's content moderation decisions, which can "result in severe human rights impacts."⁶⁴ Individualized ex post decision-making is a poor vehicle for addressing systemic harms.⁶⁵

Commentators have been especially critical of the OB's inability to review cases in which Meta removes content to comply with local law. Construed broadly, this provision would preclude the Board from considering situations in which governments censor online content or engage in disinformation campaigns—both of which have significant adverse human rights consequences.⁶⁶ The Board also lacks authority over important areas of the social media company's business, including Meta's content ranking algorithms.⁶⁷

Another set of critiques relates to the Board's constrained authority in compelling Meta to disclose relevant information. The Charter commits Meta to "provid[ing] information, in compliance with applicable legal and privacy restrictions, that is reasonably required for the board to make a decision."⁶⁸ However, the company appears to reserve to itself the determination of what information is "reasonably required," and it has rejected requests that would have helped the Board make informed decisions.⁶⁹

⁶² *Id.* art. 2.

⁶³ See, e.g., Editorial, *Weaponizing Facebook*, PHILIPPINE DAILY INQUIRER (Dec. 10, 2021, 5:08 AM), <https://opinion.inquirer.net/147373/weaponizing-facebook> [<https://perma.cc/9G34-BTEH>]; *UN Rights Experts Raise Alarm over Russia's 'Choking' Media Clampdown at Home*, UNITED NATIONS (Mar. 11, 2022), <https://news.un.org/en/story/2022/03/1113762> [<https://perma.cc/K9FB-YL6W>]; *Philippines: Events of 2020*, HUM. RTS. WATCH, <https://www.hrw.org/world-report/2021/country-chapters/philippines> [<https://perma.cc/4GGW-DWUV>]. The BSR recommended that non-Facebook and non-Instagram users "have a channel to access the Oversight Board for use if content directly or indirectly impacts them." BSR, *supra* note 39, at 5.

⁶⁴ BSR, *supra* note 39, at 12, 16; evelyn douek, *Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 808–09 (2021); see also Klonick, *supra* note 6.

⁶⁵ douek, *supra* note 20, at 569.

⁶⁶ douek, *supra* note 9, at 38–39; Parmar, *supra* note 9.

⁶⁷ douek, *supra* note 9, at 39–44; O'Kane, *supra* note 10, at 192–94.

⁶⁸ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 3, § 3.

⁶⁹ See, e.g., Former President Trump's Suspension, No. 2021-001-FB, 2021 WL 4822219, at *13 (May 5, 2021); Punjabi Concern over the RSS in India, No. 2021-003-FB, 2021 WL 4822221, at *8 (Apr. 29, 2021); Öcalan's Isolation, No. 2021-006-IG, 2021 WL 4822224, at *11–12 (July 8, 2021); see also O'Kane, *supra* note 10, at 200 ("Where Meta refuses to cooperate, the benefits of the Board and its weak-form review are limited to public awareness.").

The Charter and Bylaws also adopt a narrow conception of compliance.⁷⁰ The Bylaws explain that while “[t]he board’s resolution on each case will be binding on Facebook, . . . the policy advisory statement from the board will be considered as a recommendation.”⁷¹ Thus, only the outcome of a specific case is binding. The company has agreed to consider the feasibility of applying the Board’s decisions to “identical content with parallel context.”⁷² However, Meta need not give individual decisions wider application if it concludes that this is not feasible.⁷³ As for OB advisory statements, the Charter commits the company only to take these policy recommendations “into consideration,”⁷⁴ including by “analyzing the operational procedures required to implement the guidance, considering it in the formal policy development process of Facebook, and transparently communicating about actions taken as a result.”⁷⁵

Concerns have also been raised about Meta’s ability to circumvent or limit the Board’s powers.⁷⁶ Although noncompliance could result in reputational consequences, Meta may have an incentive to disregard at least some Board decisions.⁷⁷ If the company changes its Community Standards or Values, the OB would be obligated to apply these revised rules—at least prospectively.⁷⁸ Meta also retains significant power to amend the Bylaws.⁷⁹

⁷⁰ Dia Kayyali & Jillian C. York, *The Facebook Oversight Board Is Making Good Decisions—but Does It Matter?*, TECH POLICY PRESS (July 28, 2021), <https://techpolicy.press/the-facebook-oversight-board-is-making-good-decisions-but-does-it-matter> [<https://perma.cc/S9GU-5VN8>].

⁷¹ OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 2.3; *see also* OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 4; *id.* art. 3, § 7.3.

⁷² OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 2.3.1.

⁷³ *Id.* In its Human Rights Report for 2020–2021, Meta reports reinstating “identical content with parallel context” in four cases, noting in addition that “where the board upholds our judgment, we continue to ensure identical content with parallel context remains either up or down, in line with the board’s decision.” META, META HUMAN RIGHTS REPORT: INSIGHTS AND ACTIONS 2020–2021, at 73 (2022), https://about.fb.com/wp-content/uploads/2022/07/Meta_Human-Rights-Report-July-2022.pdf [<https://perma.cc/YN22-86PB>].

⁷⁴ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 3, § 4.

⁷⁵ *Id.* art. 4.

⁷⁶ Although Meta’s commitment to abide by the Board’s decisions on individual cases has been characterized as contractual, Gradoni, *supra* note 28, the OB’s founding documents do not provide for judicial or other remedies for the company’s noncompliance with Board decisions.

⁷⁷ douek, *supra* note 9, at 60.

⁷⁸ Di Stefano, *supra* note 36, at 113; *see also* Thomas Kadri, *How Supreme a Court?*, SLATE (Nov. 19, 2018, 1:59 PM), <https://slate.com/technology/2018/11/facebook-zuckerberg-independent-speech-content-appeals-court.html> [<https://perma.cc/U3HK-8XPH>].

⁷⁹ The company can modify the kinds of cases the OB reviews, the Board’s resources, the information the company provides, its implementation of OB decisions, as well as submission procedures and timelines. However, these changes can only be adopted after “consulting” with the

Finally, the Board's existence depends on continued funding from Meta. Meta has funded the OB robustly in the form of an initial contribution to the irrevocable Trust of \$130 million in 2019 when the Board was created, and a second contribution of \$150 million in 2022.⁸⁰ After that funding is exhausted, however, Meta will decide whether to provide additional support, although presumably the Board could also seek funding from other sources.

C. *A Better Analogy: International Human Rights Tribunals*

Many critiques of the Oversight Board and Meta's ability to circumvent its decisions compare the OB to a domestic court. Seen from this perspective, the Board is indeed inadequate. We contend, however, that a better analogy for the OB is found in international human rights tribunals. The following discussion identifies OB design elements that are similar to central features of those tribunals, and then reviews several salient differences.

1. Similarities

The Oversight Board is similar to international human rights tribunals in several important respects. First, both institutions have limited jurisdiction conferred upon them by their creators, as opposed to national courts that derive their authority from and are part of a broader legal system. The Board's Charter, much like a treaty establishing an international tribunal, defines the cases the OB can hear, who can bring them, and the legal issues it can address.⁸¹ In this respect, the Board mirrors the structure of international monitoring bodies, whose jurisdiction, access rules, and review powers are circumscribed by the consent of the states that created them, and which are vulnerable to their creators imposing new limits on that consent or withdrawing it altogether.⁸²

Board and the trustees, "provided that no amendment it makes to these sections will contradict the board's charter." OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 5, § 1. The Charter can only be amended with the approval of a majority of the individual trustees, Meta, and a majority of the Board. OVERSIGHT BOARD CHARTER, *supra* note 16, art. 6, § 1.

⁸⁰ *Securing Ongoing Funding for the Oversight Board*, OVERSIGHT BD. (July 22, 2022), <https://www.oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board> [<https://perma.cc/QDP6-ZC3J>].

⁸¹ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 4.

⁸² *E.g.*, Joost Pauwelyn & Rebecca J. Hamilton, *Exit from International Tribunals*, 9 J. INT'L DISP. SETTLEMENT 679, 690 (2018).

Critics who lament the Board's limited ability to act as a meaningful check on Meta underestimate the constraints of this consent-based design. Limited jurisdiction and authority are often necessary compromises because states (like Meta) are unwilling to delegate more expansive review powers to these bodies—at least at the outset. For example, the U.N. Human Rights Committee, created during an era when U.N. members were wary of international monitoring, was given only modest authority to review compliance with the International Covenant on Civil and Political Rights (ICCPR).⁸³ And in regional human rights systems, states could join a treaty without accepting the jurisdiction of the associated court.⁸⁴

When international review mechanisms are first established, therefore, the choice is often not between more or less authority, but between less authority and no authority at all. This perspective finds support from the fact that no other social media company has yet established a body similar to the OB. A few platforms have created bodies of external experts to advise them on content moderation policies.⁸⁵ However, these bodies have no binding authority vis-à-vis their respective companies, and they can be disbanded at any time—as recently occurred with Twitter's Trust and Safety Council.⁸⁶

Second, international human rights tribunals, like the Board (but unlike most domestic courts), have both binding and nonbinding functions. Regional human rights courts, for example, possess both contentious and advisory jurisdiction.⁸⁷ Only the former results in legally binding judgments, and the latter is often more frequently used during a tribunal's early years.⁸⁸ All of the normative output of the ten U.N. treaty bodies—general comments, country-specific concluding observations, and individual decisions—are authoritative but nonbinding. Much like

⁸³ See Louis Henkin, *The International Bill of Rights: The Universal Declaration and the Covenants*, in INTERNATIONAL ENFORCEMENT OF HUMAN RIGHTS 1, 7–9 (Rudolf Bernhardt & John Anthony Jolowicz eds., 1987); International Covenant on Civil and Political Rights, art. 40, Dec. 16, 1966, S. TREATY DOC. 95-20, 999 U.N.T.S. 171.

⁸⁴ E.g., ED BATES, THE EVOLUTION OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS: FROM ITS INCEPTION TO THE CREATION OF A PERMANENT COURT OF HUMAN RIGHTS 175–76 (2010).

⁸⁵ E.g., *Safety Partners*, TIKTOK, <https://www.tiktok.com/safety/en-us/safety-partners> [<https://perma.cc/MVA9-5Z3T>] (describing functions of the Content Advisory Council).

⁸⁶ Sheila Dang, *Twitter Dissolves Trust and Safety Council*, REUTERS (Dec. 13, 2022, 5:37 AM), <https://www.reuters.com/technology/twitter-dissolves-trust-safety-council-2022-12-13> [<https://perma.cc/5QB5-VDXR>].

⁸⁷ See, e.g., Jorge Contesse, *Judicial Interactions and Human Rights Contestations in Latin America*, 12 J. INT'L DISP. SETTLEMENT 271, 277, 283 (2021) (referring to the advisory and contentious jurisdiction of the Inter-American Court of Human Rights).

⁸⁸ *Id.* For example, in its first five years, the Inter-American Court of Human Rights exercised its contentious jurisdiction once and its advisory jurisdiction four times. See Thomas Buergenthal, *The Advisory Practice of the Inter-American Human Rights Court*, 79 AM. J. INT'L L. 1, 2 (1985).

the OB's policy guidance to Meta, states must give only "due consideration" to treaty body recommendations.⁸⁹

Third, the Oversight Board exercises fact-finding and promotional functions comparable to many tribunals. To inform its review of Meta's decision to remove or retain content, the OB meets with experts and advocacy organizations,⁹⁰ similar to U.N. treaty bodies that confer with civil society organizations when reviewing the human rights record of state parties.⁹¹ The Board also publishes annual reports that analyze how its "decisions have considered or tracked the international human rights implicated by a case," as well as "Facebook's implementation and response to board decisions and policy advisory statements."⁹² This is similar to the treaty bodies, which report each year to the U.N. General Assembly on their human rights activities.⁹³ The Board has also solicited information outside of the content review process, similar to human rights commissions that proactively investigate violations in particular situations or countries.⁹⁴

Fourth, both international human rights tribunals and the Board face overwhelming demand. The backlog of tens of thousands of cases pending before the ECtHR is the most extreme example, but the problem is systemic. Other regional courts and commissions, as well as some U.N. treaty bodies, receive far more petitions than they can review within a reasonable timeframe. Simply put, the tribunals cannot be expected to

⁸⁹ *E.g.*, Machiko Kanetake, *Giving Due Consideration: A Normative Pathway Between UN Human Rights Treaty-Monitoring Bodies and Domestic Courts*, in ENTANGLED LEGALITIES BEYOND THE STATE 133 (Nico Krisch ed., 2021).

⁹⁰ *See, e.g.*, Sudan Graphic Video, No. 2022-002-FB, 2022 Meta OB LEXIS 4, at *17 (June 13, 2022); Swedish Journalist Reporting Sexual Violence Against Minors, No. 2021-016-FB, 2022 WL 1479058, at *6 (Feb. 1, 2022).

⁹¹ *See* U.N. OFF. OF THE HIGH COMM'R FOR HUM. RTS., WORKING WITH THE UNITED NATIONS HUMAN RIGHTS PROGRAMME: A HANDBOOK FOR CIVIL SOCIETY, at 59–73, U.N. Doc. HR/PUB/06/10/Rev.1 (2008), https://digitallibrary.un.org/record/653716/files/Handbook_en.pdf?ln=en [<https://perma.cc/V77C-QSWK>] (addressing the role of civil society in treaty body monitoring processes).

⁹² OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 1, § 4.1.

⁹³ *See* U.N. OFF. OF THE HIGH COMM'R FOR HUM. RTS., *supra* note 91, at 59–73 (addressing the reporting requirements for each U.N. treaty monitoring body).

⁹⁴ *See, e.g.*, Douglass Cassel, *The Inter-American Human Rights System: A Functional Analysis*, in 1 LIBER AMICORUM: HÉCTOR FIX-ZAMUDIO 521, 527–28 (1998). In late 2021, the OB met with Frances Haugen, a whistleblower and former Meta employee who provided internal documents to Congress and the press about shortcomings in the company's content moderation policies. *Oversight Board to Meet with Frances Haugen*, OVERSIGHT BD. (Oct. 11, 2021), <https://www.oversightboard.com/news/1232363373906301-oversight-board-to-meet-with-frances-haugen> [<https://perma.cc/NVE6-A2LZ>]; Ryan Mac & Cecilia Kang, *Whistle-Blower Says Facebook 'Chooses Profits over Safety'*, N.Y. TIMES (Oct. 27, 2021), <https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html> [<https://perma.cc/SC3L-F6S9>].

achieve individualized justice given the pervasiveness and scale of human rights harms.⁹⁵ The Board, too, is unable to provide individual justice in light of the volume of Meta’s content moderation actions worldwide.⁹⁶

Fifth, international human rights tribunals, like the OB, have only limited ability to enforce their decisions. This is true not only for nonbinding policy guidance and recommendations, but also for legally binding judgments. Even for the ECtHR—widely recognized as the most active and influential human rights court—partial and delayed compliance is the norm, and outright resistance is not uncommon.⁹⁷ The compliance records of other tribunals are generally weaker. This has not, however, prevented these bodies from requesting information about implementation and following up with states in subsequent interactions⁹⁸—a practice that the Oversight Board has also adopted.⁹⁹ These interactions sometimes induce countries to take incremental steps toward compliance, a trend that can be accelerated when civil society groups rely on tribunal decisions as focal points for domestic and transnational advocacy.¹⁰⁰

Sixth, and finally, states can deploy a variety of tools to weaken or undermine international human rights tribunals, just as Meta can vis-à-vis the OB. Scholars have catalogued a range of “pushback” and “backlash” techniques, including critiquing specific judgments or doctrines, narrowing a tribunal’s delegated authority, changing the composition of its members, starving it of resources, or withdrawing from its jurisdiction.¹⁰¹ Although resistance to national courts is also common in some countries, opposition to international tribunals can, in the most extreme cases, lead to their de facto suspension.¹⁰²

⁹⁵ Henry J. Steiner, *Individual Claims in a World of Massive Violations: What Role for the Human Rights Committee?*, in *THE FUTURE OF UN HUMAN RIGHTS TREATY MONITORING* 15, 31–33 (Philip Alston & James Crawford eds., 2000).

⁹⁶ Haggart & Keller, *supra* note 11, at 8–9.

⁹⁷ See, e.g., Darren Hawkins & Wade Jacoby, *Partial Compliance: A Comparison of the European and Inter-American Courts of Human Rights*, 6 J. INT’L L. & INT’L RELS. 35, 66–76 (2010).

⁹⁸ See, e.g., Walter Kälin, *Examination of State Reports*, in *UN HUMAN RIGHTS TREATY BODIES: LAW AND LEGITIMACY* 16, 27–28 (Helen Keller & Geir Ulfstein eds., 2012) [hereinafter *UN HUMAN RIGHTS TREATY BODIES*].

⁹⁹ See generally *OVERSIGHT BOARD ANNUAL REPORT*, *supra* note 18.

¹⁰⁰ See generally KATHRYN SIKKINK, *THE JUSTICE CASCADE: HOW HUMAN RIGHTS PROSECUTIONS ARE CHANGING WORLD POLITICS* (2011); BETH A. SIMMONS, *MOBILIZING FOR HUMAN RIGHTS: INTERNATIONAL LAW IN DOMESTIC POLITICS* (2009).

¹⁰¹ Mikael Rask Madsen, Pola Cebulak & Micha Wiebusch, *Backlash Against International Courts: Explaining the Forms and Patterns of Resistance to International Courts*, 14 INT’L J.L. CONTEXT 197, 211–13 (2018).

¹⁰² See generally COURTNEY HILLEBRECHT, *SAVING THE INTERNATIONAL JUSTICE REGIME: BEYOND BACKLASH AGAINST INTERNATIONAL COURTS* (2021).

Notwithstanding these constraints and limitations, most human rights tribunals have incrementally expanded their competencies over time and developed international norms without eliciting extensive state opposition.¹⁰³ In later Sections of this Article, we discuss and suggest ways to mitigate the risk that Meta will oppose the OB for making similar moves.

2. Differences

There are also a variety of differences between the Oversight Board and human rights tribunals. We focus here on four distinctions that are likely to have significant influence on the Board's future trajectory.

The first, and most obvious, difference is that Meta is a private corporation and not a state.¹⁰⁴ Some limits on the Board's jurisdiction reflect the fact that, unlike human rights tribunals, the Board oversees a private company that can incur liability as a result of its activities around the world. States are responsible for violations of international law attributable to them, and their officials may face civil or criminal liability for such violations. However, a commercial entity like Meta also has obligations to shareholders that require it to avoid foreseeable financial or legal risks, which include retaining on its platforms content that violates domestic law. In addition, while human rights treaties are negotiated and ratified by multiple states, the Board has only a single creator. This enhances Meta's ability to ignore Board decisions and policy recommendations that, in the company's view, are too aggressive in promoting accountability or developing international norms.¹⁰⁵

A second distinction is the vast financial support that Meta has provided to the Board (\$280 million as of 2022, as noted above) in comparison to the much lesser funding that states provide to human rights tribunals. Only the ECtHR has anything close to the OB's budget (€74.5 million for 2022),¹⁰⁶ but it reviews a vastly greater number of

¹⁰³ See, e.g., Laurence R. Helfer, *Pushback Against Supervisory Systems: Lessons for the ILO from International Human Rights Institutions*, in ILO100: LAW FOR SOCIAL JUSTICE 257, 277 (George P. Politakis, Tomi Kohiyama & Thomas Lieby eds., 2019).

¹⁰⁴ See generally Benesch, *supra* note 35.

¹⁰⁵ See Tom Ginsburg, *Bounded Discretion in International Judicial Lawmaking*, 45 VA. J. INT'L L. 631, 670 (2005) (noting that the power of a tribunal to make law is correlated with the number of parties to a regime, the difficulty of amending the treaty, and the costs of exit).

¹⁰⁶ COUNCIL OF EUR., COUNCIL OF EUROPE PROGRAMME AND BUDGET 2022–2025 tbl. 1 (2021), <https://rm.coe.int/0900001680a4d5de> [<https://perma.cc/PH7B-7H4W>].

2023]

META OVERSIGHT BOARD

2255

applications each year (over 36,000 in 2021)¹⁰⁷ than does the OB (20 in 2021).¹⁰⁸ In comparison, other international human rights tribunals are chronically starved for resources.¹⁰⁹

A third difference relates to the ability to screen and select cases for review. Human rights tribunals apply admissibility criteria to determine whether they can adjudicate a petition on the merits. Among the most common is the exhaustion of domestic remedies, a rule that is analogous to requiring a Facebook user to file an internal challenge to the company's content moderation decision before bringing a case to the Board.¹¹⁰ Once the admissibility criteria have been met, however, the tribunals have no discretion to not hear a case (although they can dispose of petitions without published decisions).¹¹¹

In contrast, the OB selects which cases it reviews. The Board has prioritized "cases that had a potential to affect lots of users around the world, were critically important to public discourse or raised important questions about Meta's policies."¹¹² These included cases involving automated moderation, Meta's "dangerous individuals and organizations" standard, and press freedom.¹¹³ In addition, the Board has selected cases that "raised major implications for applying international human rights standards to moderating content at global scale," including cases "from outside the Global North."¹¹⁴

A fourth difference concerns remedies. The Oversight Board's remedial powers in individual cases are narrow: it can direct Meta to uphold or reverse the company's internal determination to retain or take

¹⁰⁷ Compare EUR. CT. OF HUM. RTS., EUROPEAN COURT OF HUMAN RIGHTS STATISTICS 2021 (2022), https://www.echr.coe.int/Documents/Stats_annual_2021_ENG.pdf [https://perma.cc/Q98X-JGYV], with *An Empirical Look at the Facebook Oversight Board*, LAWFARE, <https://www.lawfareblog.com/empirical-look-facebook-oversight-board> [https://perma.cc/NMC7-GF6B].

¹⁰⁸ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 5.

¹⁰⁹ See, e.g., Vincent O. Nmehielle, *Financing and Sustaining the African Court of Justice and Human and Peoples' Rights*, in *THE AFRICAN COURT OF JUSTICE AND HUMAN AND PEOPLES' RIGHTS IN CONTEXT: DEVELOPMENT AND CHALLENGES* 1057, 1057 (Charles C. Jalloh, Kamari M. Clarke & Vincent O. Nmehielle eds., 2019) (stating that the desire to merge the African Court on Human and Peoples' Rights and the African Court of Justice is "predicated on the increasingly diminishing resources available").

¹¹⁰ See, e.g., EUR. CT. OF HUM. RTS., PRACTICAL GUIDE ON ADMISSIBILITY CRITERIA 26–38 (2022), https://www.echr.coe.int/documents/admissibility_guide_eng.pdf [https://perma.cc/5WG5-TNT6].

¹¹¹ See, e.g., INT'L FED'N FOR HUM. RTS., ADMISSIBILITY OF COMPLAINTS BEFORE THE AFRICAN COURT: PRACTICAL GUIDE 13 (2016), <https://www.refworld.org/pd/577cd89d4.pdf> [https://perma.cc/KR8B-LBMP].

¹¹² OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 15.

¹¹³ *Id.*

¹¹⁴ *Id.* at 15, 20.

down content.¹¹⁵ The remedies awarded by the tribunals are far more capacious. They include declaring that rights have been violated, awarding compensation, and recommending that a state take concrete actions, such as releasing an individual from detention, reopening judicial proceedings, and adopting legislative or other systemic measures to remedy the structural causes of treaty violations.¹¹⁶

D. *The Impact of International Human Rights Tribunals:
Accountability Promotion and Norm Development*

For some commentators, comparing the Oversight Board to international human rights tribunals may reinforce their view that the OB is unlikely to have much of an impact. An extensive literature has critiqued the effectiveness of the tribunals on multiple grounds, including the proliferation of monitoring bodies, the fragmented jurisprudence they produce, their lack of enforcement powers, and the low levels of compliance with their decisions.¹¹⁷ Other studies, in contrast, note that the existence of multiple bodies enables advocates to strategically choose the forum that is most advantageous for seeking accountability.¹¹⁸ They also identify the tribunals' positive impact on applying international law to new contexts and enabling civil society organizations to push governments to implement human rights protections domestically.¹¹⁹

For those familiar with domestic legal systems, it may seem obvious that the effectiveness of any adjudicatory body depends on whether litigants obey its rulings. In the international arena, however, whether the parties to a dispute follow the orders and provide the remedies that a tribunal awards is only one measure of impact, and arguably not the most

¹¹⁵ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 1, § 4(4).

¹¹⁶ *See, e.g.*, DINAH SHELTON, REMEDIES IN INTERNATIONAL HUMAN RIGHTS LAW 191–238 (3d ed. 2015).

¹¹⁷ *See, e.g.*, EMILIE M. HAFNER-BURTON, MAKING HUMAN RIGHTS A REALITY 95–123 (2013) (identifying shortcomings of international human rights bodies as including inadequate resources, underused or ineffective complaints mechanisms, cumbersome reporting processes, fragmented legal norms, and weak enforcement); HURST HANNUM, RESCUING HUMAN RIGHTS: A RADICALLY MODERATE APPROACH 97–118 (2019) (discussing proliferation and fragmentation).

¹¹⁸ *See, e.g.*, GRÁINNE DE BÚRCA, REFRAMING HUMAN RIGHTS IN A TURBULENT ERA 202–03 (2021).

¹¹⁹ *See, e.g.*, Rachel Murray & Christian De Vos, *Behind the State: Domestic Mechanisms and Procedures for the Implementation of Human Rights Judgments and Decisions*, 12 J. HUM. RTS. PRAC. 22, 34–40 (2020); Cosette D. Creamer & Beth A. Simmons, *The Dynamic Impact of Periodic Review on Women's Rights*, 81 L. & CONTEMP. PROBS. 31, 32 (2018); Alejandro Anaya-Muñoz, Hector M. Nuñez & Aldo F. Ponce, *Setting the Agenda: Social Influence in the Effects of the Human Rights Committee in Latin America and Central and Eastern Europe*, 17 J. HUM. RTS. 229, 232–38 (2018).

important one.¹²⁰ Other measures of effectiveness include whether tribunal decisions deter future violations;¹²¹ influence behavior in a rights-enhancing direction;¹²² become embedded in domestic law, thus enabling violations to be remedied at home;¹²³ promote the systemic integration of international law;¹²⁴ and adapt human rights agreements to changing circumstances.¹²⁵

The wide variety of these metrics reflects the fact that international human rights tribunals engage in rights-promoting activities beyond adjudicating individual complaints. They also investigate systemic violations, review state reports, issue recommendations to improve compliance, and provide interpretive guidance in general comments and advisory opinions.¹²⁶ The precise functions vary from one institution to another. In general, however, all international human rights tribunals share two overarching objectives: (1) providing a venue to hold states accountable for legal violations, and (2) developing international norms in response to longstanding and emerging human rights issues.

The Sections that follow discuss how human rights tribunals promote accountability and develop international norms and explain how those functions have evolved over time. This sets the stage for our proposal for the Oversight Board to prioritize similar goals.

¹²⁰ Alexandra Huneus, *Compliance with Judgments and Decisions*, in THE OXFORD HANDBOOK OF INTERNATIONAL ADJUDICATION 437, 439–40 (Cesare P.R. Romano, Karen J. Alter & Yuval Shany eds., 2014) [hereinafter OXFORD HANDBOOK].

¹²¹ See generally JILLIENNE HAGLUND, REGIONAL COURTS, DOMESTIC POLITICS, AND THE STRUGGLE FOR HUMAN RIGHTS 24–75 (2020) (analyzing the conditions under which regional human rights courts deter future human rights violations).

¹²² See Karen J. Alter, Laurence R. Helfer & Mikael Rask Madsen, *International Court Authority in a Complex World*, in INTERNATIONAL COURT AUTHORITY 3, 4 (Karen J. Alter, Laurence R. Helfer & Mikael Rask Madsen eds., 2018); Theresa Squatrito, Oran R. Young, Andreas Follesdal & Geir Ulfstein, *A Framework for Evaluating the Performance of International Courts and Tribunals*, in THE PERFORMANCE OF INTERNATIONAL COURTS AND TRIBUNALS 3, 16 (Theresa Squatrito, Oran R. Young, Andreas Follesdal & Geir Ulfstein eds., 2018).

¹²³ See Laurence R. Helfer, *Redesigning the European Court of Human Rights: Embeddedness as a Deep Structural Principle of the European Human Rights Regime*, 19 EUR. J. INT'L L. 125, 133–41 (2008); Machiko Kanetake, *UN Human Rights Treaty Monitoring Bodies Before Domestic Courts*, 67 INT'L & COMPAR. L.Q. 201, 206–18 (2018).

¹²⁴ See Armin von Bogdandy & Ingo Venzke, *In Whose Name? An Investigation of International Courts' Public Authority and Its Democratic Justification*, 23 EUR. J. INT'L L. 7, 36–38 (2012).

¹²⁵ Laurence R. Helfer, *The Effectiveness of International Adjudicators*, in OXFORD HANDBOOK, *supra* note 120, at 464, 466.

¹²⁶ See, e.g., Lutz Oette, *The UN Human Rights Treaty Bodies: Impact and Future*, in INTERNATIONAL HUMAN RIGHTS INSTITUTIONS, TRIBUNALS, AND COURTS 95 (Gerd Oberleitner ed., 2018).

1. Promoting Accountability

Despite the limits that states initially placed on the powers of international human rights tribunals, these bodies have, over time, incrementally expanded their authority to adjudicate complaints and monitor implementation and compliance. These expansions have enhanced the tribunals' ability to hold states accountable for violations of human rights law. Although states have not explicitly authorized these expansions, they have generally acquiesced in the tribunals' augmented review powers, in part due to pressure from civil society organizations.¹²⁷ A few examples illustrate these developments.

The ECtHR, the oldest and most active human rights tribunal, conceives of itself not merely as the judicial arm of the European Convention on Human Rights but as the guardian of a "constitutional instrument of European public order."¹²⁸ The court has adopted numerous doctrinal innovations that reflect this elevated status. It has severed reservations that it deems incompatible with the Convention's object and purpose, treated interim measures as legally binding notwithstanding states' positions that such measures are only recommendations, and created an aggregate litigation mechanism to address systemic human rights problems.¹²⁹

U.N. treaty bodies have also capaciously construed their mandates to review state party reports, formulate general comments, and issue decisions in response to individual complaints.¹³⁰ For example, the "views" that the Human Rights Committee adopts in response to complaints from individuals are nonbinding, but the Committee expects states to follow them nonetheless. It has also taken concrete steps to monitor compliance, appointing a special rapporteur to evaluate state responses and collecting and publicizing compliance information. In these ways, the Human Rights Committee and other treaty bodies have

¹²⁷ See, e.g., Par Engstrom & Peter Low, *Mobilising the Inter-American Human Rights System: Regional Litigation and Domestic Human Rights Impact in Latin America*, in *THE INTER-AMERICAN HUMAN RIGHTS SYSTEM: IMPACT BEYOND COMPLIANCE* 23, 25–26 (Par Engstrom ed., 2019); Cosette D. Creamer & Beth A. Simmons, *The Proof Is in the Process: Self-Reporting Under International Human Rights Treaties*, 114 *AM. J. INT'L L.* 1, 36–39 (2020).

¹²⁸ *Loizidou v. Turkey*, 310 *Eur. Ct. H.R.* (ser. A) at 22 (1995).

¹²⁹ See, e.g., Iain Cameron & Frank Horn, *Reservations to the European Convention on Human Rights: The Belilos Case*, 33 *GERMAN Y.B. INT'L L.* 69, 115–16 (1990); Yves Haeck, Clara Burbano Herrera & Leo Zwaak, *Non-Compliance with a Provisional Measure Automatically Leads to a Violation of the Right of Individual Application . . . or Doesn't It?*, 4 *EUR. CONST. L. REV.* 41, 51 (2008); Markus Fyrnys, *Expanding Competences by Judicial Lawmaking: The Pilot Judgment Procedure of the European Court of Human Rights*, 12 *GERMAN L.J.* 1231, 1232–33 (2011).

¹³⁰ Laurence R. Helfer & Anne-Marie Slaughter, *Toward a Theory of Effective Supranational Adjudication*, 107 *YALE L.J.* 273, 344 (1997).

pushed “the limits of [their] authority and sometimes arguably [gone] beyond it.”¹³¹

In the Americas, the Inter-American Court of Human Rights (IACtHR) has developed a range of responses to enhance its effectiveness, including issuing periodic reports on whether states have complied with its judgments and the remedies they award.¹³² The court also created the doctrine of conventionality control, which requires all domestic governmental actors to give effect to the American Convention on Human Rights as interpreted by the IACtHR. The doctrine reflects a particularly strong view of international judicial supremacy, one in which national judges are compliance partners for the court’s binding judgments and advisory opinions.¹³³

2. Developing International Norms

In addition to enhancing their ability to hold states accountable for violations, the tribunals also contribute to the promotion and protection of human rights via interpretation. The development of international norms stems from a shared conception of human rights treaties as “living instrument[s]”¹³⁴ that “tak[e] into account the circumstances of contemporary society,”¹³⁵ as well as progressive trends at the global, regional, and national levels.¹³⁶ Such interpretations have encompassed a wide array of topics, including the right to life, inhuman and degrading treatment, discrimination against LGBT persons, conscientious objection

¹³¹ *Id.*

¹³² Jeffrey K. Staton & Alexia Romero, *Rational Remedies: The Role of Opinion Clarity in the Inter-American Human Rights System*, 63 INT’L STUD. Q. 477, 481–82 (2019).

¹³³ See generally INTER-AM. COMM’N ON HUM. RTS., COMPENDIUM ON THE OBLIGATION OF STATES TO ADAPT THEIR DOMESTIC LEGISLATION TO THE INTER-AMERICAN STANDARDS OF HUMAN RIGHTS 25–48 (2021); PABLO GONZÁLEZ-DOMÍNGUEZ, THE DOCTRINE OF CONVENTIONALITY CONTROL: BETWEEN UNIFORMITY AND LEGAL PLURALISM IN THE INTER-AMERICAN HUMAN RIGHTS SYSTEM 63–176 (2018).

¹³⁴ *Tyrer v. United Kingdom*, 26 Eur. Ct. H.R. (ser. A) at 12 (1978); see also *Magyar Helsinki Bizottság v. Hungary*, App. No. 18030/11, ¶ 3 (Nov. 8, 2016) (Sicilianos & Raimondi, JJ., concurring), <https://hudoc.echr.coe.int/eng?i=001-167828> [<https://perma.cc/AAK4-7DFZ>] (explaining that the living instrument doctrine has “spread throughout the Strasbourg case-law,” enabling the ECtHR “to adapt, over time, the text of the Convention to legal, social, ethical or scientific developments”).

¹³⁵ Comm. on the Elimination of Racial Discrimination, *General Recommendation No. 32: The Meaning and Scope of Special Measures in the International Convention on the Elimination of All Forms Racial Discrimination*, ¶ 5, U.N. Doc. CERD/C/GC/32 (Sept. 24, 2009).

¹³⁶ See, e.g., Gerald L. Neuman, *Import, Export, and Regional Consent in the Inter-American Court of Human Rights*, 19 EUR. J. INT’L L. 101, 106–07 (2008); Birgit Schlütter, *Aspects of Human Rights Interpretation by the UN Treaty Bodies*, in UN HUMAN RIGHTS TREATY BODIES, *supra* note 98, at 261, 261.

to military service, and collective bargaining by trade unions.¹³⁷ In these and other areas, the tribunals have viewed with increasing skepticism rights-restrictive laws and policies once accepted as within a state's discretion.

Two examples illustrate these trends. The first concerns how consensus emerged around an expanded definition of rape as a form of torture, a practice prohibited in many human rights treaties. As one scholar recently explained:

[W]hen the Special Rapporteur first asserted, in 1986, that rape can constitute torture, this idea was quickly taken up by other bodies. Twelve years later, the ICTY, the ICTR, the IACHR and the ECtHR had all found that the prohibition of torture and ill-treatment had been violated by acts of rape.¹³⁸

A second example focuses on the general comments of the Committee on Economic, Social and Cultural Rights. Over several decades, the Committee has developed a range of doctrines and principles—including the nonderogable minimum core; obligations to respect, protect, and fulfill; and the extension of duties to private actors—that have augmented and concretized provisions of the International Covenant on Economic, Social and Cultural Rights.¹³⁹ Individuals and civil society groups, in turn, have relied on these normative developments in litigation, especially in countries in the Global South, to progressively interpret economic and social rights in national constitutions.¹⁴⁰ Studies have documented similar examples of the international and domestic diffusion of norms developed by other U.N. treaty bodies and regional human rights courts.¹⁴¹

¹³⁷ See, e.g., Hum. Rts. Comm., *General Comment No. 36: Article 6 of the International Covenant on Civil and Political Rights, on the Right to Life*, ¶¶ 10–31, U.N. Doc. CCPR/C/GC/36 (Sept. 3, 2019); Rudolf Bernhardt, *Evolutionary Treaty Interpretation, Especially of the European Convention on Human Rights*, 42 GERMAN Y.B. INT'L L. 11, 14 (1999); ELAINE WEBSTER, DIGNITY, DEGRADING TREATMENT AND TORTURE IN HUMAN RIGHTS LAW: THE ENDS OF ARTICLE 3 OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS 28–43 (2018).

¹³⁸ CORINA HERI, RESPONSIVE HUMAN RIGHTS: VULNERABILITY, ILL-TREATMENT AND THE ECtHR 6 (2021) (footnote omitted).

¹³⁹ See, e.g., Philip Alston, *The General Comments of the UN Committee on Economic, Social and Cultural Rights*, 104 AM. SOC'Y INT'L L. PROC. 4, 5–7 (2010); Michael J. Dennis & David P. Stewart, *Justiciability of Economic, Social, and Cultural Rights: Should There Be an International Complaints Mechanism to Adjudicate the Rights to Food, Water, Housing, and Health?*, 98 AM. J. INT'L L. 462, 491–500 (2004).

¹⁴⁰ See generally Laurence R. Helfer, *Pharmaceutical Patents and the Human Right to Health: The Contested Evolution of the Transnational Legal Order on Access to Medicines*, in TRANSNATIONAL LEGAL ORDERS 311 (Terence C. Halliday & Gregory Shaffer eds., 2015).

¹⁴¹ See, e.g., Jorge Contesse, *The Rule of Advice in International Human Rights Law*, 115 AM. J. INT'L L. 367, 377–81 (2021); Creamer & Simmons, *supra* note 127, at 40–41; Laurence R. Helfer &

2023]

META OVERSIGHT BOARD

2261

As these examples reveal, normative development often begins within a single human rights treaty. Over time, however, evolutive interpretations can spread to other international human rights conventions, as well as within and across domestic legal systems. The result is a force multiplier effect that can dramatically enhance the influence and impact of human rights norms. As we later explain, a similar trend is emerging for the Oversight Board, whose normative output is already being cited within the U.N. human rights system.

II. THE OVERSIGHT BOARD'S HUMAN RIGHTS FUTURE

As we have explained, commentators have criticized the Oversight Board on numerous grounds, including its narrow jurisdiction, limited capacity to review user challenges, and inability to compel Meta to change its policies. Similar criticisms have been leveled at international human rights tribunals. Yet as the previous Section has shown, the tribunals have incrementally enhanced their authority and progressively expanded international norms to increase their influence over governments.

We now propose a range of strategies for the Oversight Board to chart a similar trajectory. We begin by explaining how the OB has already begun to hold Meta accountable for its content moderation decisions and policies. We then discuss how the Board is developing international human rights norms, in particular with respect to how these norms apply to private social media companies. We then identify strategies for the Board to continue along this trajectory, drawing where appropriate on principles and doctrines that international human rights tribunals have developed.

A. *Accountability Promotion*

1. Jurisdiction

The Oversight Board's decisions thus far reflect a fairly broad interpretation of its own jurisdiction. From the beginning, it rejected Meta's argument that an appeal is rendered moot by the company's

Erik Voeten, *International Courts as Agents of Legal Change: Evidence from LGBT Rights in Europe*, 68 INT'L ORG. 77, 91–105 (2014); Pammela Quinn Saunders, *The Integrated Enforcement of Human Rights*, 45 N.Y.U. J. INT'L L. & POL. 97, 117–70 (2012).

removal or restoration of the content at issue.¹⁴² It has also expansively exercised its authority to provide Meta with policy guidance, issuing robust recommendations in nearly every case even though this function appears to have been intended as subsidiary to its review of individual decisions.¹⁴³ Despite fears that the Board would not be able to address the ways in which Meta's recommendation systems affect human rights, the OB has capaciously interpreted its policy advisory function to recommend that the company "commission a human rights impact assessment on how its newsfeed, recommendation algorithms, and other features amplify harmful health misinformation and its impacts."¹⁴⁴

In the future, we expect that the Board will continue to face borderline cases in which it will need to determine its own jurisdiction. All international tribunals have such authority. Although the exercise of this power is mostly uncontroversial, a tribunal may encounter "Kompetenz-Kompetenz" disputes over whether to expand its jurisdictional reach.¹⁴⁵

For example, the OB may need to clarify its jurisdiction with respect to content removed by Meta to comply with domestic law. Some commentators claim that this exception precludes the Board from "consider[ing] cases that human rights activists care most deeply about: state censorship of content, particularly political speech, in flagrant violation of international human rights standards on freedom of expression."¹⁴⁶ In our view, however, the OB's founding documents provide leeway for the Board to hear cases raising these issues. At a minimum, the Board has the authority to determine whether the criteria for this exception have been met.

The Charter provides that nothing in the OB's founding documents "shall be interpreted in a manner that would result in a violation of law by Facebook" or any of its associated entities.¹⁴⁷ The Bylaws, in turn,

¹⁴² Breast Cancer Symptoms and Nudity, No. 2020-004-IG, 2021 WL 4822215, at *8-9 (Jan. 28, 2021); see also OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 27 ("[T]he Board has the authority to review cases from users even when Meta chooses to later rectify its mistake and restore the content."); Wong & Floridi, *supra* note 3.

¹⁴³ O'Kane, *supra* note 10, at 187.

¹⁴⁴ Oversight Bd., *Removal of COVID-19 Misinformation*, No. PAO-2022-01 (Apr. 20, 2023), <https://oversightboard.com/attachment/547865527461223> [<https://perma.cc/E726-EN98>].

¹⁴⁵ The term "Kompetenz-Kompetenz" originates in German constitutional law. Applied to international adjudication, the term refers to a tribunal's power to determine its own jurisdiction. *Constitutional Court v. Peru, Competence, Judgment*, Inter-Am. Ct. H.R. (ser. C) No. 55, ¶ 31 (Sept. 24, 1999); see also Gulati, *supra* note 14, at 9-10 (arguing that the Oversight Board possesses "competence-competence").

¹⁴⁶ Parmar, *supra* note 9.

¹⁴⁷ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 7 ("The board will not purport to enforce local law.").

2023]

META OVERSIGHT BOARD

2263

identify three categories of cases that “will not be eligible for the board to review”:

- Where the underlying content has already been blocked, following the receipt of a valid report of illegality, and not removed for a Community Standards violation;

- Where the underlying content is criminally unlawful in a jurisdiction with a connection to the content (such as the jurisdiction of the posting party and/or the reporting party) and where a board decision to allow the content on the platform could lead to criminal liability for Facebook, Facebook employees, the administration, or the board's members; or
- Where the underlying content is unlawful in a jurisdiction with a connection to the content (such as the jurisdiction of the posting party and/or the reporting party) and where a board decision to allow the content on the platform could lead to adverse governmental action against Facebook, Facebook employees, the administration, or the board's members.¹⁴⁸

As an initial matter, the three exceptions to the OB's jurisdiction in the Bylaws are broader than those in the Charter, which only requires the Board to refrain from decisions that would result in liability for Meta. However, the exceptions in the Bylaws are more constrained than they may initially appear. The first exception excludes review only when content has been blocked pursuant to a "valid report of illegality." The practices of both Meta and the Board indicate that this exception only prevents review of content that is removed directly pursuant to a government request. When the OB has asked the company about the existence of such a request, Meta has provided a substantive response.¹⁴⁹ In addition, Meta and the Board agree that this exception does not include situations in which the content was flagged by a government entity but was later removed for a Community Standards violation.¹⁵⁰

Perhaps the strongest indication that the "valid report of illegality" exception will be interpreted narrowly appears in a 2022 decision of the Board, referred by Meta, in which the company removed a drill music video that had been brought to its attention by an internet referral unit in the United Kingdom.¹⁵¹ Such law enforcement offices, which use internet providers' terms of service to seek removal of content that the

¹⁴⁸ OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 1.2.2.

¹⁴⁹ Video After Nigeria Church Attack, No. 2022-011-IG, 2022 Meta OB LEXIS 16, at *10 (Dec. 14, 2022) ("In response to questions from the Board, Meta confirmed that the Nigerian government did not contact Meta regarding the attack or request that the content be taken down.").

¹⁵⁰ Öcalan's Isolation, No. 2021-006-IG, 2021 WL 4822224, at *12 (July 8, 2021). The Board expressed concern that users are not informed when the state is involved in content removals, and that government flagging is not reported in the company's transparency reports. These concerns led the Board to conclude that Meta was not respecting the right to a remedy under the U.N. Guiding Principles on Business and Human Rights. *Id.* at *11–12.

¹⁵¹ Oversight Board (@OversightBoard), TWITTER (July 26, 2022, 8:05AM), <https://twitter.com/OversightBoard/status/1551901459847020544?s=20&t=OlqmEetdMpmr8xM85NMBgA> [<https://perma.cc/JE4C-7EQL>].

2023]

META OVERSIGHT BOARD

2265

government deems objectionable,¹⁵² raise serious human rights concerns.¹⁵³ The Board found there was insufficient evidence of a credible veiled threat and issued several recommendations for how Meta should respond to such requests in the future.¹⁵⁴

If the Board hears cases implicating other exceptions to its jurisdiction, we believe it should determine for itself whether the exceptions are applicable. In addition to demonstrating its independence and rejecting Meta's ability to exclude cases from Board review simply by asserting the existence of an exception, this approach would have normative benefits. For example, it would provide the Board with an opportunity to address an unresolved tension in the U.N. Guiding Principles on Business and Human Rights ("U.N. Guiding Principles" or "UNGPs"): the obligation of businesses to follow local law while, at the same time, respecting human rights.¹⁵⁵ In determining what constitutes a "valid report of illegality," for example, the OB might draw on nonbinding standards such as the Manila Principles on Intermediary Liability¹⁵⁶ or the Global Network Initiative (GNI) Implementation Guidelines.¹⁵⁷ More audaciously, the Board might conclude that Meta

¹⁵² Similar referral units have been launched by Europol as well as the Netherlands, Belgium, Italy, Germany, and Israel. See, e.g., Jason Pielemeier & Chris Sheehy, *Understanding the Human Rights Risks Associated with Internet Referral Units*, MEDIUM (Feb. 25, 2019), <https://medium.com/global-network-initiative-collection/understanding-the-human-rights-risks-associated-with-internet-referral-units-by-jason-pielemeier-b0b3feeb95c9> [<https://perma.cc/2SBZ-53YG>]; Daphne Keller, *When Platforms Do the State's Bidding, Who Is Accountable? Not the Government, Says Israel's Supreme Court*, LAWFARE (Feb. 7, 2022, 1:01 PM), <https://www.lawfareblog.com/when-platforms-do-states-bidding-who-accountable-not-government-says-israels-supreme-court> [<https://perma.cc/GZZ2-MNQ9>].

¹⁵³ Molly K. Land, *Against Privatized Censorship: Proposals for Responsible Delegation*, 60 VA. J. INT'L L. 363, 401–02 (2020).

¹⁵⁴ UK Drill Music, No. 2022-007-IG, 2022 Meta OB LEXIS 6, at *6–7, *58–62 (Nov. 22, 2022). The OB recommended that Meta make public how much content is removed pursuant to content policies following a government request; standardize its process for receiving such requests; request a more detailed evidentiary basis from the referring agency; and "regularly review the data on its content moderation decisions prompted by state actor content review requests to assess for any systemic biases." *Id.* at *58–62.

¹⁵⁵ John Ruggie (Special Representative of the Sec'y Gen. on the Issue of Hum. Rts. & Transnat'l Corps. & Other Bus. Enters.), *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, U.N. Doc. A/HRC/17/31, ¶ 23(a)–(b) (Mar. 21, 2011) [hereinafter *U.N. Guiding Principles*].

¹⁵⁶ *Manila Principles on Intermediary Liability*, MANILA PRINCIPLES, <https://manilaprinciples.org/index.html> [<https://perma.cc/M8G2-4JD4>].

¹⁵⁷ *Implementation Guidelines*, GLOB. NETWORK INITIATIVE, <https://globalnetworkinitiative.org/implementation-guidelines> [<https://perma.cc/3SJA-X7UZ>]. Meta's Corporate Human Rights Policy, announced on March 16, 2021, commits the company to "respecting human rights as set out in the United Nations Guiding Principles on Business and

may disregard domestic criminal or civil legislation that an international tribunal has found to violate human rights, as is the case with several national media laws in Africa.¹⁵⁸

We acknowledge, however, that compliance with local law is a sensitive issue that many transnational corporations face. The Board should thus proceed carefully, recognizing that Meta has compliance expertise and a range of business interests affected by domestic legal requirements. However, nothing in the Charter or Bylaws precludes the OB from asking the company to justify its invocation of an exception or provide information about how it responds to government requests and claims of illegality. Nor do the Board's founding documents prevent it from providing the policy guidance on the overarching issue that the exceptions raise: how Meta reconciles its human rights responsibilities and compliance with local law.

2. Standards of Review

Neither the Charter nor the Bylaws address what standard of review the Oversight Board should apply. This Section identifies a range of circumstances in which the OB will confront whether to give any deference to the decisions, actions, policies, or interpretations of Meta or of governments. We draw upon doctrines and approaches developed by international human rights tribunals to identify factors that the Board should consider when determining appropriate standards of review.

Like the OB's founding documents, human rights treaties that create international tribunals are silent on standards of review.¹⁵⁹ However, the tribunals' accountability promotion and norm development functions weigh against giving any deference to states regarding the meaning of international law. Consistent with this perspective, the tribunals have

Human Rights," as well as the GNI Principles and Implementation Guidelines. *Corporate Human Rights Policy*, META (Mar. 16, 2021), <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf> [<https://perma.cc/4RNL-VT85>].

¹⁵⁸ See Fed'n of Afr. Journalists v. Gambia, Judgment No. ECW/CCJ/JUD/04/18, ECOWAS Community Court of Justice [ECOWAS Cmty. Ct. of Just.] (Feb. 13, 2018), http://www.courtecowas.org/wp-content/uploads/2019/02/ECW_CCJ_JUD_04_18.pdf [<https://perma.cc/2LEP-G7X4>] (concluding that media laws on sedition, false news, and criminal defamation violated the right to freedom of expression); Media Council of Tanz. v. Att'y Gen. of Tanz., Ref. No. 2/2017, East African Court of Justice [E. Afr. Ct. of Just.] (Mar. 28, 2019), <https://www.eacj.org/wp-content/uploads/2019/03/Referene-No.2-of-2017.pdf> [<https://perma.cc/5FE3-PRLP>] (concluding that Tanzania's Media Service Act, which gives government agencies broad power to censor and limit the offline and online content, contravened the right to freedom of expression).

¹⁵⁹ Yuval Shany, *Toward a General Margin of Appreciation Doctrine in International Law?*, 16 EUR. J. INT'L L. 907, 911 (2005).

generally applied de novo review when interpreting human rights agreements and customary law.¹⁶⁰

At the same time, the tribunals have identified circumstances in which deference to governments may be appropriate. This is most famously—and controversially—reflected in the ECtHR’s margin of appreciation doctrine.¹⁶¹ Other human rights tribunals at least implicitly recognize that a degree of deference is appropriate in some contexts.¹⁶² This includes situations where “national authorities enjoy comparative institutional advantages”¹⁶³—for example, in interpreting and applying domestic law, evaluating local conditions, or making certain factual findings—or where international rules permit those authorities “to strike a balance between competing rights and interests” or make “a choice between competing social values.”¹⁶⁴

The accountability promotion and norm development functions that the Oversight Board shares with international tribunals favor giving no deference to Meta’s interpretation of international human rights law when the company sets content moderation policies or reviews user posts. The social media company has no special expertise in this regard, nor are there institutional or normative reasons for deference. Accordingly, the Board should exercise de novo review.

Whether the Board should give any deference to Meta’s assessment of whether a post violates its Values or Community Standards is a closer question. The caselaw to date suggests that the OB continues to wrestle with this question. In *Video After Nigeria Church Attack*, for example, the panel was divided over the question of whether Meta made the right judgment call in removing a video showing the aftermath of a terrorist attack that included shooting-related hashtags. Meta (and the minority of the OB) argued that the hashtags could be viewed as sadistic, glorifying violence, and mocking, and therefore should be removed under three different Community Standards. The majority disagreed. Apparently giving no deference to Meta’s determination, the majority concluded that

¹⁶⁰ See generally DEFERENCE IN INTERNATIONAL COURTS AND TRIBUNALS: STANDARD OF REVIEW AND MARGIN OF APPRECIATION (Lukasz Gruszczynski & Wouter Werner eds., 2014).

¹⁶¹ See generally ANDREW LEGG, THE MARGIN OF APPRECIATION IN INTERNATIONAL HUMAN RIGHTS LAW: DEFERENCE AND PROPORTIONALITY (2012).

¹⁶² Although only the ECtHR has explicitly adopted the margin of appreciation doctrine, other tribunals have developed a “practice of employing substitute, MoA-like approaches.” Yuval Shany, *All Roads Lead to Strasbourg?: Application of the Margin of Appreciation Doctrine by the European Court of Human Rights and the UN Human Rights Committee*, 9 J. INT’L DISP. SETTLEMENT 180, 181 (2018); see also Nino Tsereteli, *Emerging Doctrine of Deference of the Inter-American Court of Human Rights?*, 20 INT’L J. HUM. RTS. 1097, 1106 (2016).

¹⁶³ Shany, *supra* note 159, at 913.

¹⁶⁴ Shany, *supra* note 162, at 183.

the hashtags were only designed to raise awareness among airsoft and firearm enthusiasts and thus did not violate the Community Standards.¹⁶⁵

Is de novo review of Community Standards appropriate? Meta retains the power under the Charter to modify its internal norms.¹⁶⁶ This arguably includes the lesser power to interpret the Values and Community Standards. If the company adopts a general position regarding the meaning of those internal norms, there may be structural or functional reasons for deferring to its position. In assessing how much deference, if any, is appropriate in that context, relevant factors for the Board may include whether the interpretation is based on experience with moderating specific types of online speech, whether Meta follows that interpretation consistently, and whether the company has considered the potential disadvantages of its position.¹⁶⁷ In contrast, the arguments for discretion are weaker when reviewing the decisions of front-line moderators applying Community Standards or Values to specific user posts.

The political and social context of user posts is another setting in which standard of review issues may arise. The Board appears to apply something akin to de novo review of Meta's assessment of how contextual factors affect the likelihood of real-world harm. In *Protest in India Against France*, for example, a majority of the OB "did not find Facebook's contextual rationale in relation to possible violence in India in this particular case [to be] compelling."¹⁶⁸ In contrast, a minority of the panel would have "defer[red] to Facebook's determination that the post presented an unacceptable risk of promoting violence,"¹⁶⁹ in part, because the company "relied upon a third party partner assessment" and "consulted regional and linguistic experts."¹⁷⁰

The majority's position seems compelling. The Board has the resources to engage deeply with stakeholders and develop a nuanced understanding of context in the small number of cases it selects for review. In contrast, while Meta has broad experience working with local governments and evaluating demands for compliance with local law, its

¹⁶⁵ Video After Nigeria Church Attack, No. 2022-011-IG, 2022 Meta OB LEXIS 16, at *22–26 (Dec. 14, 2022).

¹⁶⁶ See OVERSIGHT BOARD CHARTER, *supra* note 16 (setting out the powers of the Board).

¹⁶⁷ Cf. *Animal Sci. Prods., Inc. v. Hebei Welcome Pharm. Co.*, 138 S. Ct. 1865, 1873 (2018) (identifying circumstances that govern the deference that federal courts should give to a foreign state's views about the meaning of its own laws).

¹⁶⁸ *Protest in India Against France*, No. 2020-007-FB, 2021 WL 4822218, at *5 (Feb. 12, 2021).

¹⁶⁹ *Id.*

¹⁷⁰ *Id.*

2023]

META OVERSIGHT BOARD

2269

understanding of the real-world human rights impact of the speech on its platforms is more limited.¹⁷¹

Another standard of review issue relates to the Board's evaluation of government decisions and actions. The OB obviously lacks expertise in all of the local laws that may be implicated by its oversight functions. However, it has substantial resources to commission experts to analyze the domestic laws and the decisions of public actors that are relevant to an appeal, a capacity that is important given the risk that governments may not be fully forthcoming. The Board can also initiate interactions with public agencies and officials. In *UK Drill Music*, for example, the Board received a comment from the London Metropolitan Police and it filed a Freedom of Information Act request to compel the law enforcement body "to provide information about its policies and practice on making requests to social media and streaming companies to review and/or remove content."¹⁷² Taken together, these considerations suggest that the OB should give weight to a government's understanding of its own laws while preserving the possibility of verifying how the law applies in particular cases.

3. Applicable Law

A third set of strategic issues relates to the legal rules and norms that the OB applies. As previously explained, the Charter directs the OB to "pay particular attention to the impact of removing content in light of human rights norms protecting free expression" when evaluating cases.¹⁷³ Many Board decisions analyze freedom of expression in depth, as we discuss in detail below. However, the OB has not shied away from addressing other human rights. For example, it has applied the prohibition on gender and racial discrimination as well as the rights to privacy and to an effective remedy.¹⁷⁴ The OB has also considered the obligations of the International Covenant on Economic, Social and Cultural Rights, such as the right to health and to participate in culture,

¹⁷¹ See, e.g., Juncal Montero Regules, *The Facebook Oversight Board and 'Context'*, VERFASSUNGSBLOG (Feb. 16, 2021), <https://verfassungsblog.de/fob-context> [<https://perma.cc/9A8S-TZPA>].

¹⁷² UK Drill Music, No. 2022-007-IG, 2022 Meta OB LEXIS 6, at *28–29 (Nov. 22, 2022).

¹⁷³ OVERSIGHT BOARD CHARTER, *supra* note 18, art. 2, § 2.

¹⁷⁴ See, e.g., Breast Cancer Symptoms and Nudity, No. 2020-004-IG, 2021 WL 4822215, at *5, *8 (Jan. 28, 2021); Depiction of Zwarte Piet, No. 2021-002-FB, 2021 WL 4822220, at *4, *8 (Apr. 13, 2021). In addition, Meta itself has requested the Board's advisory guidance on the privacy implications of its content policies. Oversight Bd., *Sharing Private Residential Information*, No. PAO-2021-01, ¶ 1 (Feb. 8, 2021), <https://oversightboard.com/attachment/628701241555465> [<https://perma.cc/FTJ7-F8AY>].

as well as the Convention on the Rights of the Child, among other treaties.¹⁷⁵

Board decisions have cited to and applied a wide range of international human rights sources. These include international conventions (most often the ICCPR and the International Convention on the Elimination of All Forms of Racial Discrimination), general comments and decisions of treaty bodies (in particular those of the Human Rights Committee), reports submitted to the U.N. Human Rights Council (especially those of the Special Rapporteur on the Right to Freedom of Opinion and Expression), and nonbinding human rights standards (notably, the UNGPs and the hate speech guidelines of the Rabat Plan of Action).¹⁷⁶ By engaging with these sources and citing its own prior decisions applying them, the OB is developing a well-reasoned, consistent, and coherent body of human rights jurisprudence.¹⁷⁷

There are several accountability-promoting advantages to the Board's commitment to "plac[e] human rights at the heart of [its] work."¹⁷⁸ The first is stability. Meta can change its Values and Community Standards by applying its ordinary corporate decision-making processes. In contrast, the Charter cannot be amended without majority approval of the trustees and agreement of both the Board and Facebook.¹⁷⁹ The Charter's reference to international human rights law thus serves as an external check on Meta. A second benefit is contemporary relevance. As previously explained, the interpretation of human rights law evolves over time. The Board's consistent practice of applying not only treaty texts but also the normative output of

¹⁷⁵ See, e.g., *Asking for Adderall*, No. 2021-015-FB, 2022 WL 1479057, at *7 (Feb. 1, 2022); *Breast Cancer Symptoms and Nudity*, 2021 WL 4822215, at *7; see also *Wampum Belt*, No. 2021-012-FB, 2021 WL 5992273, at *9 (Dec. 9, 2021). But see Michael Wu, *The Facebook Oversight Board's Decision on COVID Misinformation and Conspicuous Omission of the Right to Health*, UCI L. INT'L JUST. CLINIC (Apr. 30, 2021), <https://ijclinic.law.uci.edu/2021/04/30/the-facebook-oversight-boards-decision-on-covid-misinformation-and-the-conspicuous-omission-of-the-right-to-health> [<https://perma.cc/7U9H-VB2V>] (noting the failure of the Board to cite the right to health).

¹⁷⁶ See, e.g., Joan Barata, *The Decisions of the Oversight Board from the Perspective of International Human Rights Law*, COLUMBIA U. GLOB. FREEDOM OF EXPRESSION 10 (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/4TVT-3ZNK>] (reviewing the international human rights instruments cited in OB decisions, including the Rabat Plan of Action).

¹⁷⁷ See OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 47 ("Applying international human rights to content moderation can help . . . provide a consistent, global framework for holding social media companies to account."); see, e.g., *Reclaiming Arabic Words*, No. 2022-003-IG, 2022 Meta OB LEXIS 5, at *8-9 (June 13, 2022) (applying prior OB decisions when analyzing Meta's human rights responsibilities).

¹⁷⁸ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 43.

¹⁷⁹ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 6, § 1.

2023]

META OVERSIGHT BOARD

2271

international tribunals and independent experts thus helps to ensure that it will continue to apply standards that respond to real-world problems and controversies.¹⁸⁰

A more challenging issue concerns potential conflicts among the sources that the Board applies. The OB's founding documents do not indicate a hierarchy among Community Standards, Values, and international human rights law. Most Board decisions to date apply the three sources independently of one another, giving each equal weight and finding no conflict among them.¹⁸¹ However, the Board has indicated that it will "prioritize human rights" when that law conflicts with Meta's internal rules.¹⁸² In *Pro-Navalny Protests in Russia*, the Board overturned the company's decision to remove a post by a supporter of the Russian opposition politician, finding that the removal was consistent with Community Standards but inconsistent with international human rights law and with Meta's Values (which incorporate human rights).¹⁸³ The OB has also made recommendations on how the company can "bring [its] Community Standard[s] in line with Meta's values and international human rights standards."¹⁸⁴

If the Board faces other clashes between the company's internal norms and human rights law, it can apply interpretive tools that international human rights tribunals have developed to mitigate these conflicts. For example, the tribunals presume that international organizations and treaty drafters do "not intend to impose any obligation on member States to breach fundamental principles of human rights."¹⁸⁵ They also apply a *pro homine* principle to construe textual ambiguities in favor of individual rights and against state sovereignty.¹⁸⁶ And under the principle of *lex specialis*, the tribunals interpret general legal obligations in light of more specific commitments.¹⁸⁷ We expect that there will be few

¹⁸⁰ Klonick notes that the Board has started to discuss the extent to which it can adopt an evolutionary approach to interpretation. Klonick, *supra* note 6.

¹⁸¹ See, e.g., Knin Cartoon, No. 2022-001-FB, 2022 Meta OB LEXIS 3, at *22–38 (June 17, 2022); South Africa Slurs, No. 2021-011-FB, 2021 WL 4822229, at *6–11 (Sept. 28, 2021); Armenians in Azerbaijan, No. 2020-003-FB, 2021 WL 4822214, at *5–8 (Jan. 28, 2021).

¹⁸² OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 46.

¹⁸³ Pro-Navalny Protests in Russia, No. 2021-004-FB, 2021 WL 4822222, at *7–11 (May 26, 2021).

¹⁸⁴ Ayahuasca Brew, No. 2021-013-IG, 2021 WL 5992274, at *8 (Dec. 9, 2021).

¹⁸⁵ Al-Dulimi v. Switzerland, App. No. 5809/08, ¶ 140 (June 21, 2016), <https://hudoc.echr.coe.int/eng?i=001-164515> [<https://perma.cc/BZ3Q-TC9H>].

¹⁸⁶ See, e.g., Yota Negishi, *The Pro Homine Principle's Role in Regulating the Relationship Between Conventionality Control and Constitutionality Control*, 28 EUR. J. INT'L L. 457, 467–73 (2017).

¹⁸⁷ See, e.g., Lautsi v. Italy, 2011-III Eur. Ct. H.R. 61, 91.

inconsistencies among sources that the OB cannot resolve by applying these or other interpretive doctrines.

4. Evidentiary Issues

Once the Oversight Board has exercised jurisdiction and has identified the applicable legal norms, it must determine whether it has an adequate factual record to render a decision. The Board regularly receives information from a variety of sources. It often commissions expert reports on the socio-political and cultural context of the content at issue and receives public comments, including from U.N. Special Rapporteurs and advocacy organizations.¹⁸⁸

Meta has thus far actively participated in cases reviewed by the Board. It has provided information about its decisions to remove or retain user posts as well as its content moderation and internal appeals processes.¹⁸⁹ In several cases, however, the OB has expressed frustration with the limited information that Meta has provided in response to its requests.¹⁹⁰ The company's disclosures and omissions have also revealed inconsistencies between Meta's publicly available policies and the rules applied by its content moderators, resulting in Board recommendations to close these gaps and provide greater transparency.¹⁹¹

The OB has yet to address when information in Meta's possession is "reasonably required for the board to make a decision," nor has it considered situations in which the company has withheld information "in compliance with applicable legal and privacy restrictions."¹⁹² The Charter does not define these key phrases. To fill in the gaps, the Board may seek guidance from international human rights tribunals that have faced similar issues.

¹⁸⁸ See, e.g., Armenians in Azerbaijan, No. 2020-003-FB, 2021 WL 4822214, at *3-4 (Jan. 28, 2021); see also OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 52.

¹⁸⁹ The Bylaws list the types of information that Meta discloses in all cases as well as additional information that the Board may request. OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 2.2.2. The Bylaws describe the information that Meta can withhold, which includes information that "is not reasonably required for decision-making in accordance with the intent of the charter, is not technically feasible to provide, is covered by attorney/client privilege, and/or cannot or should not be provided because of legal, privacy, safety, or data protection restrictions or concerns." *Id.*

¹⁹⁰ Swedish Journalist Reporting Sexual Violence Against Minors, No. 2021-016-FB, 2022 WL 1479058, at *9 (Feb. 1, 2022); Punjabi Concern over the RSS in India, No. 2021-003-FB, 2021 WL 4822221, at *8 (Apr. 29, 2021).

¹⁹¹ See Nazi Quote, No. 2020-005-FB, 2021 WL 4822216, at *5 (Jan. 28, 2021).

¹⁹² OVERSIGHT BOARD CHARTER, *supra* note 16, art. 3, § 3.

States are expected to cooperate fully in international proceedings.¹⁹³ Significantly, the tribunal, not the government, determines whether the state has produced all relevant information or conducted an adequate investigation.¹⁹⁴ This is so even where the government raises confidentiality or security objections. In such cases, the tribunal undertakes an independent verification as to whether or not there had actually “existed reasonable and solid grounds for treating the documents in question as secret or confidential.”¹⁹⁵ When states fail to live up to these obligations, the tribunals have responded by creating evidentiary presumptions to incentivize disclosure. Once an applicant makes out a prima facie case, the failure to produce information plausibly within the state’s control or to conduct an adequate investigation will result in a determination that the applicant’s plausible allegations have been proven.¹⁹⁶ The Oversight Board should follow this approach if it confronts similar evidentiary disputes with Meta, including whether confidentiality or security concerns justify withholding relevant information.

5. Compliance

As previously explained, the Charter adopts a narrow definition of compliance: OB decisions are legally binding on Meta only in each specific case. The company has committed to implement the Board’s directives to restore or remove content within seven days, unless doing so “could violate the law,” and to publicize its responses.¹⁹⁷ Meta must also analyze “whether it is technically and operationally feasible to apply the board’s decision” to “identical content with parallel context.”¹⁹⁸ With regard to policy guidance, the Board’s Charter and Bylaws commit Meta to “considering it in the formal policy development process” and providing a public response within sixty days.¹⁹⁹

¹⁹³ Janowiec v. Russia, 2013-V Eur. Ct. H.R. 203, 272–73.

¹⁹⁴ *Id.*

¹⁹⁵ *Id.* at 271.

¹⁹⁶ *See, e.g.,* Velásquez-Rodríguez v. Honduras, Merits, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 4, ¶¶ 176–88 (July 29, 1988).

¹⁹⁷ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 4; OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 2.3.1.

¹⁹⁸ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 4.

¹⁹⁹ *Id.*; OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 2, § 2.3.2. O’Kane observes that these provisions in the Bylaws “have created a feedback loop that has proved relatively successful in compelling Meta to implement certain recommendations and provide clarity on how it creates and implements its policies.” O’Kane, *supra* note 10, at 190.

The OB has applied these provisions to expand the influence of its binding decisions and nonbinding policy recommendations. Most notably, it has created institutional structures to monitor implementation, including a Case Implementation and Monitoring Team, an Implementation Working Group, and an Implementation Committee.²⁰⁰ These bodies enable the Board to press Meta on unresolved compliance issues.²⁰¹ In addition, the OB is structuring its decisions in ways that facilitate tracking of Meta's implementation.²⁰² For example, the Board numbers its policy recommendations and asks the company to address each of them separately,²⁰³ specifies the actions that would constitute implementation of its recommendations,²⁰⁴ and highlights issues of ongoing noncompliance.²⁰⁵ The Board also asks Meta to provide information to facilitate disclosures that aid its accountability efforts.²⁰⁶

Like many human rights tribunals, the Board is using its reporting function—its quarterly and annual reports—to publicize Meta's compliance record. It praises the company for its progress, identifies areas where Meta has “misrepresented or misunderstood” the Board's decisions, and tracks the company's implementation.²⁰⁷ It has also developed criteria for measuring Meta's compliance and adopted a

²⁰⁰ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 36, 60. The Board has explained that the creation of these bodies represents “a clear choice to place implementation on par with our organization's most critical functions.” *Id.* at 60.

²⁰¹ For example, the Implementation Committee has been urging “Meta to provide greater transparency about its processes for identifying and taking enforcement action on pieces of content that are both identical to those featured in our decisions and presented in a parallel context.” *Id.* at 36.

²⁰² *Id.* at 57.

²⁰³ *See, e.g.*, Punjabi Concern over the RSS in India, No. 2021-003-FB, 2021 WL 4822221, at *9 (Apr. 29, 2021); “Two Buttons” Meme, No. 2021-005-FB, 2021 WL 4822223, at *12 (May 20, 2021).

²⁰⁴ *See, e.g.*, Swedish Journalist Reporting Sexual Violence Against Minors, No. 2021-016-FB, 2022 WL 1479058, at *10–11 (Feb. 1, 2022); *Sharing Private Residential Information* (No. PAO-2021-01), *supra* note 174, at 12–15.

²⁰⁵ In *Ayahuasca Brew*, for example, the Board discussed Meta's continued failure to explain to users that its content policies also apply to Instagram, noting that the Board “does not understand why Meta is unable to immediately provide users with a greater degree of transparency by updating language in the Guidelines.” *Ayahuasca Brew*, No. 2021-013-IG, 2021 WL 5992274, at *10 (Dec. 9, 2021).

²⁰⁶ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 48 (“By asking specific questions and including the details in our final decision, we can hold Meta to account and provide researchers with new information about how the company works.”).

²⁰⁷ *See, e.g., id.* (“In our first year, Meta has generally been responsive to the Board's inquiries, answering the vast majority (86%) of our questions.”); *id.* at 49 (detailing Board questions Meta did not answer); *id.* at 57 (noting that the OB was using its “quarterly transparency reports to note instances where Meta had misrepresented or misunderstood our recommendations”). *See generally* Klonick, *supra* note 7, at 2464 (noting that the requirement that Meta communicate about its responses to nonbinding recommendations “creates a weak-form review over Facebook that could potentially become powerful by creating a public system of accountability”).

2023]

META OVERSIGHT BOARD

2275

grading system to indicate whether Meta is fully responsive, partially responsive, or nonresponsive in relation to each of its decisions and policy recommendations.²⁰⁸

The Board is also pushing the company to define its own rules in the first instance. In *Former President Trump's Suspension*, for example, the OB refused to pass judgment on Meta's indefinite suspension of the former U.S. President because the company had not followed its policies regarding account suspension.²⁰⁹ Requiring Meta to "apply and justify a defined penalty"²¹⁰ prior to review incentivizes the company to articulate and comply with its own policies and safeguards the Board's independent supervisory role. The Board also appears to be using its policy advisory authority to push the company to generate information that researchers and activists can use to evaluate Meta's human rights impacts.²¹¹

The transparency and dialogue engendered by the Board's approach serves several functions. Identifying areas of compliance increases the reputational benefits to the company of following OB decisions and recommendations. Conversely, calling out inconsistencies and highlighting areas for improvement raise the costs of noncompliance and provide information that civil society groups can use to advocate for reforms. These consequences are likely to increase as the Board issues decisions clarifying how international human rights law applies to private social media companies, an issue to which we now turn.

B. *Norm Development*

The Oversight Board has the potential to make important contributions to developing international human rights norms, particularly as they apply to business entities in the digital sphere. This Section explains why the OB is well positioned to develop these norms, reviews indications that it is already doing so in its caselaw, and suggests how the Board can influence the activities of other human rights bodies

²⁰⁸ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 59 ("In assessing each response, we asked three questions: (1) Does it address all parts of our recommendation? (2) Does it provide a commitment to action? (3) Does it provide a timeline? Where Meta met one of these criteria we deemed it to have provided a 'somewhat comprehensive,' and where it met at least two we deemed it to have provided a 'comprehensive' response.").

²⁰⁹ *Former President Trump's Suspension*, No. 2021-001-FB, 2021 WL 4822219, at *20–21 (May 5, 2021).

²¹⁰ *Id.* at *21. Meta has since reinstated former President Trump's account. *Oversight Board Response to Meta's Announcement on Former President Trump*, OVERSIGHT BD. (Jan. 25, 2023), <https://www.oversightboard.com/news/6274494105946760-oversight-board-response-to-meta-s-announcement-on-former-president-trump> [<https://perma.cc/8PF7-3T9W>].

²¹¹ *Removal of COVID-19 Misinformation* (No. PAO-2022-01), *supra* note 144.

as well as government efforts to regulate corporations and social media platforms.

1. The Oversight Board as Norm Generator

There is an urgent need to develop human rights norms, especially freedom of expression, as applied to private actors in the online environment. Commentators have made persuasive arguments that such norms should apply to social media platforms. However, there are several areas in which further development is needed.²¹²

Although private companies have human rights responsibilities under the U.N. Guiding Principles, these are different from government obligations regarding freedom of expression.²¹³ Speech restrictions imposed by platforms do not carry the coercive power that accompanies state action, and the Board has cited this as justification for upholding Meta's removal of content that a state could not restrict in similar circumstances.²¹⁴ Platforms also have different interests than governments: they pursue commercial objectives rather than societal welfare, and their own speech interests also play a role. It remains unclear how these interests should be accommodated within the human rights framework governing free expression. The online environment also affects the proportionality analysis in crucial ways.²¹⁵ For example, the large volume of information collected by social media companies may provide an evidentiary basis for crafting speech norms that more precisely target communications that cause real-world harms.²¹⁶ These and other differences have important practical consequences for the development of human rights norms relating to online expression.

A few national courts have evaluated social media content moderation under constitutional or fundamental rights standards.²¹⁷ The

²¹² *Rep. of the Special Rapporteur, supra* note 35, ¶¶ 41–43.

²¹³ *See generally* Sander, *supra* note 35; Aswad, *The Future of Freedom of Expression Online, supra* note 35; Benesch, *supra* note 35; Molly K. Land, *Regulating Private Harms Online: Content Regulation Under Human Rights Law*, in *HUMAN RIGHTS IN THE AGE OF PLATFORMS* 285 (Rikke Frank Jørgensen ed., 2019).

²¹⁴ Depiction of Zwarte Piet, No. 2021-002-FB, 2021 WL 4822220, at *9 (Apr. 13, 2021).

²¹⁵ *See* Sander, *supra* note 35, at 977–79.

²¹⁶ Richard Ashby Wilson & Molly K. Land, *Hate Speech on Social Media: Content Moderation in Context*, 52 *CONN. L. REV.* 1029, 1069–75 (2021).

²¹⁷ *See, e.g., Rodriguez v. Google Inc.*, COLUM. U. GLOB. FREEDOM OF EXPRESSION, <https://globalfreedomofexpression.columbia.edu/cases/rodriguez-v-google-inc> [<https://perma.cc/C5K7-RF3Z>]; *Shirin R.K. v. State of Kerala*, COLUM. U. GLOB. FREEDOM OF EXPRESSION, <https://globalfreedomofexpression.columbia.edu/cases/shirin-r-k-v-state-of-kerala> [<https://perma.cc/VYH8-V6YT>]. We are grateful to Agustina del Campo for pointing us to national court decisions reviewing social media companies' content moderation decisions.

2023]

META OVERSIGHT BOARD

2277

Oversight Board is well-placed to expand upon these developments by examining how human rights apply to non-state actors in the digital realm across a range of specific contexts. The OB's structure facilitates case-specific adjudication: it has sufficient resources to gather relevant information, and its members are professionally and geographically diverse.²¹⁸ The Board's ability to issue nonbinding recommendations also enables it to provide more far-reaching guidance than would be possible in a single binding decision.

2. Norm Development in OB Caselaw

The Board's thirty-nine decisions (as of May 2023)²¹⁹ have tackled a range of important and unsettled human rights issues. This Section provides an overview of OB caselaw addressing four such issues: corporate respect for freedom of expression, Meta's due diligence obligations, the cumulative harms of hate speech, and the importance of context in content moderation.²²⁰

The Board has identified several ways in which freedom of expression principles apply differently to corporations than to states. In upholding the removal of slurs or other hateful content, for example, the OB has emphasized that a private company has broader discretion to restrict speech than a state would have in a similar context, as long as the company articulates the basis for its restriction and ensures that its determinations are necessary and proportionate.²²¹ In particular, the Board has indicated that removing speech from the platform is a less severe sanction than the civil or criminal penalties applied by states and that Meta has different obligations, in part, due to the complexities of content moderation at scale.²²² The Board has also reasoned that social media companies have heightened responsibilities when they hold a dominant position in a market or when they are the primary source of

²¹⁸ Nonetheless, the Board's composition has been criticized as insufficiently diverse with regard to certain regions (the Middle East, North Africa, and Southeast Asia) and with respect to certain areas of expertise (human rights law and content moderation). *See, e.g.*, Jenny Domino, *Why Facebook's Oversight Board Is Not Diverse Enough*, JUST SEC. (May 21, 2020), <https://www.justsecurity.org/70301/why-facebooks-oversight-board-is-not-diverse-enough> [<https://perma.cc/XC9G-D8FZ>].

²¹⁹ *Decision*, *supra* note 17.

²²⁰ The Board has recognized the application of human rights law to a private company, as well as the consideration of context, as two of its major contributions to date. OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 23.

²²¹ *Armenians in Azerbaijan*, No. 2020-003-FB, 2021 WL 4822214, at *6 (Jan. 28, 2021); *Knin Cartoon*, No. 2022-001-FB, 2022 Meta OB LEXIS 3, at *31-32, *34-35, *38 (June 17, 2022).

²²² *Armenians in Azerbaijan*, 2021 WL 4822214, at *7; *Depiction of Zwarte Piet*, No. 2021-002-FB, 2021 WL 4822220, at *9 (Apr. 13, 2021).

news in a particular location.²²³ Finally, it has recommended that Meta develop tools to allow users to customize their online experience as a way to respond more narrowly to harmful content.²²⁴

Board decisions also elaborate on Meta's due diligence responsibilities under the U.N. Guiding Principles. The UNGPs direct business enterprises to assess and mitigate the human rights harms that they cause, contribute to, or that are "directly linked" to their operations.²²⁵ The OB has held that Meta's due diligence obligations are heightened in areas of conflict,²²⁶ as well as in situations of increased risk of harm,²²⁷ and when national protections of human rights, particularly freedom of expression, are weak.²²⁸ In such situations, the consequences of error—either of failing to remove harmful content or of removing benign content—are high, and the company must evaluate close cases carefully rather than relying on default rules.²²⁹ The Board has also provided insights on the kinds of risks that Meta should consider,²³⁰ as well as the company's obligations with respect to automated decision-making and machine learning, technical design, transparency reporting, and content moderation processes and procedures.²³¹ The OB has also emphasized that automatic closure of appeals to Meta may deprive users of remedies.²³²

Although commentators have raised concerns about the Board's ability to consider cumulative harms while reviewing individual cases, the

²²³ Punjabi Concern over the RSS in India, No. 2021-003-FB, 2021 WL 4822221, at *7 (Apr. 29, 2021); Shared Al Jazeera Post, No. 2021-009-FB, 2021 WL 4822227, at *8 (Sept. 14, 2021). For academic research making the same argument about dominant platforms, see generally EMILY B. LAIDLAW, REGULATING SPEECH IN CYBERSPACE: GATEKEEPERS, HUMAN RIGHTS AND CORPORATE RESPONSIBILITY (2015).

²²⁴ Russian Poem, No. 2022-008-FB, 2022 Meta OB LEXIS 7, at *36 (Nov. 16, 2022) (recommending that Meta develop tools to allow users to determine whether to see sensitive content).

²²⁵ *U.N. Guiding Principles*, *supra* note 155, ¶ 17(a).

²²⁶ *Shared Al Jazeera Post*, 2021 WL 4822227, at *10; Alleged Crimes in Raya Kobo, No. 2021-014-FB, 2021 WL 5992283, at *11 (Dec. 14, 2021).

²²⁷ Myanmar Bot, No. 2021-007-FB, 2021 WL 4822225, at *8 (Aug. 11, 2021).

²²⁸ Öcalan's Isolation, No. 2021-006-IG, 2021 WL 4822224, at *11 (July 8, 2021).

²²⁹ *Myanmar Bot*, 2021 WL 4822225, at *8–9.

²³⁰ *See, e.g.*, Punjabi Concern over the RSS in India, No. 2021-003-FB, 2021 WL 4822221, at *7–8 (Apr. 29, 2021); Pro-Navalny Protests in Russia, No. 2021-004-FB, 2021 WL 4822222, at *10 (May 26, 2021); Reclaiming Arabic Words, No. 2022-003-IG, 2022 Meta OB LEXIS 5, at *29–31 (June 13, 2022).

²³¹ *See, e.g.*, Oversight Bd., *Meta's Cross-Check Program*, No. PAO-2021-02 (Dec. 6, 2022), <https://oversightboard.com/attachment/440576264909311> [<https://perma.cc/6VJK-92CX>]; Breast Cancer Symptoms and Nudity, No. 2020-004-IG, 2021 WL 4822215, at *9 (Jan. 28, 2021); *Shared Al Jazeera Post*, 2021 WL 4822227, at *11; Asking for Adderall®, No. 2021-015-FB, 2022 WL 1479057, at *8, *10 (Feb. 1, 2022).

²³² *See* Iran Protest Slogan, No. 2022-013-FB, 2023 Meta OB LEXIS 1, at *55–56 (Jan. 9, 2023).

OB has discussed such harms in several hate speech decisions. In *Depiction of Zwarte Piet*, for example, the Board upheld Meta's removal of a post containing a blackface character on the grounds that "moderating content to address the cumulative harms of hate speech, even where the expression does not directly incite violence or discrimination, can be consistent with Facebook's human rights responsibilities in certain circumstances."²³³ The majority emphasized not only the psychological impact of repeated negative stereotypes but also the way in which the accumulation of such stereotypes "creates an environment where acts of violence are more likely to be tolerated and reproduce discrimination in a society."²³⁴

Perhaps the most significant development in the Board's jurisprudence, however, is its insistence that Meta do more to consider the context of speech, despite the complexities of doing so at scale.²³⁵ Although a number of recent innovations have improved Meta's ability to tailor the internal guidance it provides to front-line moderators,²³⁶ the volume of content that these moderators review precludes them from considering posts in context or exercising discretion in applying this guidance.²³⁷ The company assesses the broader socio-political context only when a challenged post is elevated to the policy staff level,²³⁸

²³³ *Depiction of Zwarte Piet*, No. 2021-002-FB, 2021 WL 4822220, at *9 (Apr. 13, 2021).

²³⁴ *Id.*; see also *Alleged Crimes in Raya Kobo*, No. 2021-014-FB, 2021 WL 5992283, at *11 (Dec. 14, 2021) (noting that the "cumulative impact" of speech "can amount to causation through a 'gradual build-up of effect'"); *Knin Cartoon*, No. 2022-001-FB, 2022 Meta OB LEXIS 3, at *38 (June 17, 2022) (finding that removal of an implicit slur was appropriate given the "cumulative effects" of such content).

²³⁵ See Robyn Caplan, *Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches*, DATA & SOC'Y, https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf [<https://perma.cc/727T-RG94>].

²³⁶ See, e.g., *Crisis Policy Protocol*, META (Jan. 25, 2023), <https://transparency.fb.com/policies/improving/crisis-policy-protocol> [<https://perma.cc/6PP7-FBGU>]; *Sri Lanka Pharmaceuticals*, No. 2022-014-FB, 2023 Meta OB LEXIS 3, at *18 (Mar. 9, 2023) (noting that use of the "spirit of the policy" exception constituted a proportional response in light of local conditions); *Iran Protest Slogan*, 2023 Meta OB LEXIS 1, at *55 (noting with approval Meta's efforts to move toward "higher strike-to-penalty thresholds").

²³⁷ See, e.g., *Pro-Navalny Protests in Russia*, No. 2021-004-FB, 2021 WL 4822222, at *7 (May 26, 2021) ("Enforcing the Community Standard appears limited to determining whether a single term is a negative character claim and whether it has been reported by the user targeted by the claim. There is no assessment of the wider context or conversation."); see also *Knin Cartoon*, 2022 Meta OB LEXIS 3, at *27–28 (explaining that Meta instructs moderators not to consider a poster's intent, purportedly to facilitate objective and consistent review).

²³⁸ See *Iran Protest Slogan*, 2023 Meta OB LEXIS 1, at *23 ("Meta has previously informed the Board that the kind of contextual analysis its policy teams can conduct to reach decisions on-escalation is not available to moderators at-scale, who must follow internal guidance."); *Pro-Navalny Protests in Russia*, 2021 WL 4822222, at *7–8; see also Wilson & Land, *supra* note 216, at 1060.

although its cross-check program increases the availability of human review for certain types of content.²³⁹

In case after case, the Board has been pushing back, emphasizing that “[c]ontext is key” and criticizing content moderation practices and policies that ignore contextual factors.²⁴⁰ In *Pro-Navalny Protests in Russia*, for example, a supporter of the Russian opposition politician who had been attacked online called out the attacker as a “cowardly bot.”²⁴¹ Meta removed this response under its anti-bullying policy. The Board did not mince words in overturning the removal: “Facebook’s blunt and decontextualized approach can disproportionately restrict freedom of expression.”²⁴²

The Board has also highlighted a range of contextual cues that Meta should consider in assessing the harms of speech. These factors include whether a post relates to a situation of armed conflict,²⁴³ how widely it is disseminated,²⁴⁴ the response of other users to the post,²⁴⁵ whether threats or calls for exclusion of a particular group are implicit or explicit,²⁴⁶ the location of the user,²⁴⁷ the timing of the post with respect to offline events,²⁴⁸ the language of the post,²⁴⁹ whether the post incorporated a meme,²⁵⁰ whether the content is “used in protest contexts or other crisis situations where the role of government is a topic of political debate,”²⁵¹ the existence of “government campaigns of disinformation against

²³⁹ See, e.g., *Meta’s Cross-Check Program* (No. PAO-2021-02), *supra* note 231, ¶¶ 17–57 (discussing “Early Response Secondary Review” and “General Secondary Review”); Mention of the Taliban in News Reporting, No. 2022-005-FB, 2022 Meta OB LEXIS 8, at *2, *9–10, *18–19 (Sept. 15, 2022) (discussing the “High Impact False Positive Override” system); Tigray Communication Affairs Bureau, No. 2022-006-FB, 2022 Meta OB LEXIS 10, at *2, *8 (Oct. 4, 2022) (discussing “Integrity Product Operations Centres”).

²⁴⁰ *Armenians in Azerbaijan*, No. 2020-003-FB, 2021 WL 4822214, at *7 (Jan. 28, 2021); see also *Pro-Navalny Protests in Russia*, 2021 WL 4822222, at *10; Former President Trump’s Suspension, No. 2021-001-FB, 2021 WL 4822219, at *19, *22 (May 5, 2021).

²⁴¹ *Pro-Navalny Protests in Russia*, 2021 WL 4822222, at *3.

²⁴² *Id.* at *7.

²⁴³ *Armenians in Azerbaijan*, 2021 WL 4822214, at *7; see also *Russian Poem*, No. 2022-008-FB, 2022 Meta OB LEXIS 7, at *32–33 (Nov. 16, 2022) (encouraging Meta to consider how its policy on incitement applies to speech that incites lawful violence—i.e., against combatants in the context of an armed conflict).

²⁴⁴ *Armenians in Azerbaijan*, 2021 WL 4822214, at *7; see also *Shared Al Jazeera Post*, No. 2021-009-FB, 2021 WL 4822227, at *10 (Sept. 14, 2021).

²⁴⁵ *Nazi Quote*, No. 2020-005-FB, 2021 WL 4822216, at *6 (Jan. 28, 2021).

²⁴⁶ *Knin Cartoon*, No. 2022-001-FB, 2022 Meta OB LEXIS 3, at *22–24 (June 17, 2022).

²⁴⁷ *Nazi Quote*, 2021 WL 4822216, at *6.

²⁴⁸ *Id.*

²⁴⁹ *Pro-Navalny Protests in Russia*, No. 2021-004-FB, 2021 WL 4822222, at *7 (May 26, 2021).

²⁵⁰ “Two Buttons” Meme, No. 2021-005-FB, 2021 WL 4822223, at *7–8 (May 20, 2021).

²⁵¹ *Iran Protest Slogan*, No. 2022-013-FB, 2023 Meta OB LEXIS 1, at *54 (Jan. 9, 2023).

opponents and their supporters” in a particular country,²⁵² the history of the region,²⁵³ the identity or reputation of the poster,²⁵⁴ the relative power dynamics between the involved parties,²⁵⁵ and inferences about the user’s intent from the language of the post.²⁵⁶

The disagreements between the Board and Meta about the extent to which the company considers context are significant. Meta’s failure to account for context has been at the heart of key debates over its human rights record, from its insufficient Burmese-speaking reviewers during the genocide in Myanmar to its initial across-the-board ban on content depicting female nipples.²⁵⁷ Although the company has invested in additional language capacity, built greater complexity into its rules, created a process to provide additional human review of potential false positives,²⁵⁸ and developed regionally specific lists of slurs,²⁵⁹ its front-line content moderation remains focused on the post itself. The Board—while acknowledging the challenges of context for such a vast platform—has not budged in its insistence that Meta take greater account of external considerations.²⁶⁰ Whether Meta will continue to revise its processes to give meaningful attention to context could be one of the most important future tests of the OB’s legitimacy.

3. The Impact of Norm Generation on Human Rights

The previous Section described how the Oversight Board is addressing a range of unsettled legal issues that arise at the intersection of human rights and content moderation. This Section considers the mechanisms by which these normative developments can affect human rights more broadly. First, OB decisions can provide focal points for civil society groups to mobilize for rights-enhancing legal and policy reforms.

²⁵² *Pro-Navalny Protests in Russia*, 2021 WL 4822222, at *10.

²⁵³ Knin Cartoon, No. 2022-001-FB, 2022 Meta OB LEXIS 3, at *36–37 (June 17, 2022).

²⁵⁴ *Id.* at *36; see also Mention of the Taliban in News Reporting, No. 2022-005-FB, 2022 Meta OB LEXIS 8, at *30, *32, *35, *37, *39 (Sept. 15, 2022).

²⁵⁵ See, e.g., *Pro-Navalny Protests in Russia*, 2021 WL 4822222, at *8; Wampum Belt, No. 2021-012-FB, 2021 WL 5992273, at *10–11 (Dec. 9, 2021).

²⁵⁶ *Pro-Navalny Protests in Russia*, 2021 WL 4822222, at *7–8.

²⁵⁷ Steve Stecklow, *Hatebook: Inside Facebook’s Myanmar Operation*, REUTERS (Aug. 15, 2018, 3:00 PM), <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate> [<https://perma.cc/MUK8-PKDC>]; Maya Rhodan, *Facebook Lifts Ban on Exposed Nipples in Breastfeeding Pictures*, TIME (June 13, 2014, 8:47 AM), <https://time.com/2869849/facebook-breastfeeding-nipples> [<https://perma.cc/HP7E-P5CR>].

²⁵⁸ This is Meta’s cross-check program. See generally *Meta’s Cross-Check Program* (No. PAO-2021-02), *supra* note 231.

²⁵⁹ Armenians in Azerbaijan, No. 2020-003-FB, 2021 WL 4822214, at *4 (Jan. 28, 2021).

²⁶⁰ See Sander, *supra* note 35, at 979–80.

Second, the Board's caselaw enables it to engage in a dialogue with other actors who promote the human rights responsibilities of internet platforms. Third, the norms developed by the OB may influence efforts to regulate social media companies in national and regional legislation.

a. Norms as Resources for Mobilization

The relationship between international human rights tribunals and civil society is “interactive and mutually constitutive.”²⁶¹ The tribunals (particularly in light of their limited funding) rely on civil society groups to identify, file, and support the cases, reports, and other interventions that provide opportunities to promote accountability and develop international norms. These organizations, in turn, mobilize around favorable tribunal decisions, incorporating them into broader rights-claiming strategies in domestic and transnational venues and pressuring governments to reform restrictive laws and policies in line with the tribunals' progressive interpretations of human rights law.²⁶²

The Oversight Board receives input from civil society organizations, and the norms it develops are likely to be important resources for those organizations. A wide range of individuals and groups have weighed in on every OB case, from the U.N. Special Rapporteur on Minority Issues to academics and advocacy groups.²⁶³ However, third parties—both non-Meta users as well as civil society and digital rights advocates—cannot appeal to the Board and do not provide input on case selection.²⁶⁴ We also found no evidence that civil society groups are coordinating with Meta users to “tee up” cases for OB review.

The Board's greater resources mean that it will need to rely less on civil society for information and support than human rights tribunals—ostensibly an advantage, but one that will limit the input the Board may receive about where its attention is most needed. Less reliance on civil society may mean these groups will have fewer opportunities to direct the

²⁶¹ Rachel A. Cichowski, *The European Court of Human Rights, Amicus Curiae, and Violence Against Women*, 50 *LAW & SOC'Y REV.* 890, 893 (2016).

²⁶² There is a vast literature on how civil society groups mobilize around international human rights tribunals. See generally *RIGHTS AND COURTS IN PURSUIT OF SOCIAL CHANGE: LEGAL MOBILISATION IN THE MULTI-LEVEL EUROPEAN SYSTEM* (Dia Anagnostou ed., 2014); *CIVIL SOCIETY, INTERNATIONAL COURTS AND COMPLIANCE BODIES* (Tullio Treves et al. eds., 2005); SIDNEY TARROW, *THE NEW TRANSNATIONAL ACTIVISM* (2005).

²⁶³ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 44. Notably, 96% of public comments come from the United States and Canada. This large percentage is due to the fact that 97% of all comments it has received were in connection with the Trump case, in which 9,986 interventions were filed. *Id.* at 51. Interventions in other decisions ranged from six to thirty, with an average of about fifteen per case. See *An Empirical Look at the Facebook Oversight Board*, *supra* note 107 (tracking the number of public comments received in connection with each Board case).

²⁶⁴ See OVERSIGHT BOARD CHARTER, *supra* note 16, art. 2, § 1.

2023]

META OVERSIGHT BOARD

2283

Board to issues of concern and cases that would be valuable to their advocacy efforts. We thus recommend that the Board's Engagement Team establish additional channels of communication to civil society groups to help inform the OB's case selection strategy.

b. Dialogue Promotion

Norm generation can also influence human rights by promoting a dialogue between rights-holders and duty-bearers.²⁶⁵ Observers of the Oversight Board have expressed similar hopes for dialogue among Meta, its users, and other actors affected by its content moderation policies. Evelyn Douek, for example, argues that “[b]ecause the substantive rules for platform standards will almost always be a matter of reasonable disagreement and contestation, legitimacy for content moderation systems needs to be earned not by unilaterally issuing ‘correct’ rules but by channeling debate about what those rules should be.”²⁶⁶

Thus far, the Oversight Board's external engagement has been robust—even more so than the public outreach by human rights tribunals. The OB publicizes the appeals it accepts and solicits broad input on these cases using a simple and accessible template.²⁶⁷ The Board also has an Engagement Team that communicates with external audiences.²⁶⁸ It holds “office hours” to discuss how to submit comments regarding the cases it has accepted,²⁶⁹ organizes stakeholder meetings focused on important issues,²⁷⁰ and shares information about the sources it relies on in its decisions.²⁷¹

The Board also actively engages with experts in deciding cases. In the *Swedish Journalist Reporting Sexual Violence Against Minors*, for example, the OB relied heavily on input from a virtual roundtable with seven advocacy organizations working on child sexual exploitation.²⁷² Similarly, in *Reclaiming Arabic Words*, “members of the Board held informative and enriching discussions with organizations that work on

²⁶⁵ See Steiner, *supra* note 95, at 31.

²⁶⁶ Evelyn Douek, *The Limits of International Law in Content Moderation*, 6 U.C. IRVINE J. INT'L TRANSNAT'L & COMPAR. L. 37, 46 (2021).

²⁶⁷ See, e.g., Email Newsletter, Oversight Bd., Across the Board (July 29, 2022) [hereinafter Across the Board] (on file with authors).

²⁶⁸ Email Newsletter, Oversight Bd., Oversight Board Announcements and Office Hours Invites (July 26, 2022) (on file with authors).

²⁶⁹ *Id.*; OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 65.

²⁷⁰ Across the Board, *supra* note 267 (reporting on monthly stakeholders' roundtable on content moderation in Africa).

²⁷¹ See, e.g., Swedish Journalist Reporting Sexual Violence Against Minors, No. 2021-016-FB, 2022 WL 1479058, at *6 (Feb. 1, 2022); Reclaiming Arabic Words, No. 2022-003-IG, 2022 Meta OB LEXIS 5, at *15–16 (June 13, 2022).

²⁷² *Swedish Journalist Reporting Sexual Violence Against Minors*, 2022 WL 1479058, at *6.

freedom of expression and the rights of LGBTQIA+ people, including Arabic speakers.”²⁷³ The OB has not made explicit the extent to which it relies on these interventions, but several of its conclusions track information submitted by third parties.²⁷⁴

However, the Board has yet to evaluate the extent to which these efforts may privilege certain stakeholders or voices over others. Just as Meta’s outreach is directed by and filtered through its staff, the Oversight Board’s staff act as gatekeepers who influence from whom the Board hears.²⁷⁵ The addition of a “user advocate,” as recommended by BSR, would help to improve outreach, especially to users who are members of vulnerable communities.²⁷⁶ The inability of civil society organizations to file appeals also hampers the extent to which these groups can proactively initiate interactions with the Board.

In addition, there is one important constituency not currently included in this dialogue: individuals who do not have Facebook or Instagram accounts.²⁷⁷ Non-users can be deeply affected by the content moderation policies of these platforms, and denying them the opportunity for input undermines the Board’s legitimacy. Until the Charter is revised to permit appeals by non-users or otherwise formalize their involvement, the Board should consider affirmatively seeking input on case selection from advocacy organizations on issues that may not generate appeals from users or Meta.

The Board’s contribution to dialogue would also benefit from more discussion of disagreements among its members. Regional human rights courts and U.N. treaty bodies allow their members to file, under their own names, concurring or dissenting opinions.²⁷⁸ Scholars have identified numerous benefits of such separate opinions, such as improving deliberation and the quality of a tribunal’s reasoning; enhancing judicial independence and transparency; and providing information to other

²⁷³ *Reclaiming Arabic Words*, 2022 Meta OB LEXIS 5, at *15.

²⁷⁴ See, e.g., Pro-Navalny Protests in Russia, No. 2021-004-FB, 2021 WL 4822222, at *6 (May 26, 2021); Shared Al Jazeera Post, No. 2021-009-FB, 2021 WL 4822227, at *7 (Sept. 14, 2021); *Swedish Journalist Reporting Sexual Violence Against Minors*, 2022 WL 1479058, at *6.

²⁷⁵ Arun, *supra* note 12, at 252.

²⁷⁶ BSR, PROGRESS REPORT: HUMAN RIGHTS AND THE OVERSIGHT BOARD 12 (2020), https://www.bsr.org/reports/BSR_Facebook_Oversight_Board_Report_on_Progress_Dec_2020.pdf [<https://perma.cc/U6DB-J84N>].

²⁷⁷ See BSR, *supra* note 39, at 33; Alice Doyle, *Out of “Site,” Out of Mind: The Facebook Oversight Board’s Exclusion of Non-User Rightsholders*, UCIL INT’L JUST. CLINIC (Mar. 5, 2021), <https://ijclinic.law.uci.edu/2021/03/05/out-of-site-out-of-mind> [<https://perma.cc/V55A-9LBV>].

²⁷⁸ See Jeffrey L. Dunoff & Mark A. Pollack, *The Road Not Taken: Comparative International Judicial Dissent*, 116 AM. J. INT’L L. 340, 366 (2022).

tribunals, future litigants, and the public about alternative future trajectories of the law.²⁷⁹

The Board's Bylaws provide that OB decisions should "include any concurring or dissenting viewpoints, if the panel cannot reach consensus."²⁸⁰ Currently, such viewpoints are summarized as part of the written decisions issued by the five-member panel of the Board that hears each case. Although the summaries of minority opinions have been getting longer in recent cases, they are still generally less developed than the majority opinions, and they do not indicate the number of OB members that support the minority position.²⁸¹ Allowing 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.

c. Norm Diffusion

The norms that the Board develops in its caselaw may also have a significant impact on the wider international human rights system.²⁸² Prior to the Board's creation, a few commentators highlighted the normative potential of social media platforms applying human rights standards. Evelyn Mary Aswad (who later became a member and then co-chair of the Board) observed that if companies applied international human rights law to their content moderation decisions, "there could be an active fountain of new 'jurisprudence' involving the ICCPR's speech protections, which could influence the direction of international freedom of expression rights."²⁸³ Sejal Parmar also queried how OB decisions would affect the work of activists, U.N. bodies, national courts, and international freedom of expression standards.²⁸⁴

Our review of the over three dozen OB cases through May 2023 reveals that the Board is already influencing the development of human

²⁷⁹ See, e.g., *id.* at 345–48; Laurence R. Helfer & Erik Voeten, *Walking Back Human Rights in Europe?*, 31 EUR. J. INT'L L. 797, 806 (2020).

²⁸⁰ OVERSIGHT BOARD BYLAWS, *supra* note 48, art. 1, § 3.1.7.

²⁸¹ See, e.g., *Depiction of Zwarte Piet*, No. 2021-002-FB, 2021 WL 4822220, at *12 (Apr. 13, 2021); *Video After Nigeria Church Attack*, No. 2022-011-IG, 2022 Meta OB LEXIS 16, at *27–28, *30 (Dec. 14, 2022).

²⁸² The Board's choices regarding its interpretive methodology may affect the extent to which other international institutions take up its decisions. For example, the OB relies heavily on secondary resources such as the reports of U.N. human rights mandate holders rather than more traditional principles of treaty interpretation. See Andreas Kulick, *Corporations as Interpreters and Adjudicators of International Human Rights—Meta's Oversight Board and Beyond*, 22 LAW & PRAC. INT'L CTS. & TRIBUNALS (forthcoming 2023) (manuscript at 11–14) (on file with authors).

²⁸³ Aswad, *The Future of Freedom of Expression Online*, *supra* note 35, at 64.

²⁸⁴ Parmar, *supra* note 9.

rights norms at the global, regional, and national levels.²⁸⁵ In the United Nations, for example, a report of the High Commissioner for Human Rights highlighted the Board's application of the Rabat Plan of Action, a nonbinding standard for hate speech that the High Commissioner had adopted a decade earlier.²⁸⁶ The Special Rapporteur on Minority Issues reviewed several early OB decisions and praised the Board's "overarching commitment to protecting vulnerable and marginalised groups."²⁸⁷ And the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression highlighted the Board's potential to offer "valuable lessons" for efforts to remedy "wrongful actions taken on the basis of disinformation or misinformation."²⁸⁸

The Board's normative output is also relevant to U.N. treaty bodies, especially the Human Rights Committee, whose General Comment No. 34 on freedoms of opinion and expression is frequently cited to and applied by the OB.²⁸⁹ In addition, the Committee on the Elimination of Discrimination Against Women and the Committee on the Elimination of Racial Discrimination may be interested in the Board's interpretations of the rights of women and racial minorities. Several OB decisions have

²⁸⁵ For a comprehensive discussion of the Board's engagement with the U.N. human rights system, see Martin Fertmann & Anna Sophia Tiedeke, *Exploring the In-Between: Mapping Interactions Between International Human Rights Institutions and the Meta Oversight Board* (2022) (unpublished manuscript) (on file with authors). At the regional level, Board decisions may be taken up by human rights courts and commissions that have extensive jurisprudence on freedom of expression. For examples of cases in which the ECtHR has found violations of the freedom of expression in connection with online activity, see Press Release, Eur. Ct. Hum. Rts., Factsheet—Access to Internet and Freedom to Receive and Impart Information and Ideas (Sept. 2022), https://www.echr.coe.int/Documents/FS_Access_Internet_ENG.pdf [<https://perma.cc/C632-9ZAS>].

²⁸⁶ U.N. OFF. OF THE HIGH COMM'R FOR HUM. RTS., UNITED NATIONS HUMAN RIGHTS REPORT 2021, at 58, U.N. Doc. HRC/NONE/2022/1 (2022), <https://www2.ohchr.org/english/OHCHRreport2021/documents/OHCHRreport2021.pdf> [<https://perma.cc/9HTC-6KQQ>]; U.N. High Comm'r for Hum. Rts., *Annual Report of the United Nations High Commissioner for Human Rights, Addendum: Report of the United Nations High Commissioner for Human Rights on the Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred*, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013).

²⁸⁷ U.N. Special Rapporteur on Minority Issues, Public Comment by UN 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/C49K-HXVP>].

²⁸⁸ Irene Khan (Special Rapporteur on the Promotion & Prot. of the Right to Freedom of Op. & Expression), *Disinformation and Freedom of Opinion and Expression*, ¶¶ 72–73, U.N. Doc. A/HRC/47/25 (Apr. 13, 2021).

²⁸⁹ See, e.g., Depiction of Zwarte Piet, No. 2021-002-FB, 2021 WL 4822220, at *8 (Apr. 13, 2021) ("The UN Human Rights Committee has made clear the protection of Article 19 extends to expression that may be considered 'deeply offensive.'" (quoting Hum. Rts. Comm., *General Comment No. 34: Article 19: Freedoms of Opinion and Expression*, ¶ 11, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011))).

opined on how certain manifestations of online violence can reflect discrimination and structural inequality.²⁹⁰ In *Gender Identity and Nudity*, the Board relied on decisions of the Human Rights Committee as well as the Convention on the Elimination of All Forms of Discrimination Against Women and the UNGPs to conclude that Meta's Adult Nudity and Sexual Activity Community Standard resulted in "disparate opportunities for expression being made available for women, trans, and gender non-binary people on its platforms."²⁹¹

Board decisions applying the UNGPs to social media companies may be especially useful to the Working Group on Business and Human Rights, created in 2011 to focus on the implementation of these principles.²⁹² Another U.N. venue is the Open-Ended Intergovernmental Working Group on Transnational Corporations and Other Business Enterprises with Respect to Human Rights, which is elaborating a binding legal instrument on the human rights responsibilities of private companies.²⁹³ The OB caselaw could, for example, inform the drafting of treaty provisions on the due diligence requirements of digital platforms.

Finally, OB decisions may also influence the efforts of supranational and national lawmakers to regulate internet platforms. In late 2022, the European Union (EU) approved the Digital Services Act (DSA), which imposes a comprehensive set of new legal obligations on social media companies and authorizes the EU, its member states, and platform users to enforce them.²⁹⁴ With regard to content moderation, the DSA mandates that digital service providers expeditiously delete posts that are illegal under national or EU law, while also empowering users to challenge removal decisions within the companies, in out-of-court dispute settlement proceedings, and by filing suits in court.²⁹⁵ Although precisely how these review processes will function will not be resolved until the DSA is transposed into national law, commentators have already

²⁹⁰ See, e.g., *id.* at *9–12; Swedish Journalist Reporting Sexual Violence Against Minors, No. 2021-016-FB, 2022 WL 1479058, at *9–10 (Feb. 1, 2022).

²⁹¹ See *Gender Identity and Nudity*, Nos. 2022-009-IG, 2022-010-IG, 2023 Meta OB LEXIS 2, at *45 (Jan. 17, 2023).

²⁹² See *Working Group on Business and Human Rights*, U.N. OFF. OF THE HIGH COMM'R FOR HUM. RTS., <https://www.ohchr.org/en/special-procedures/wg-business> [<https://perma.cc/PT7Q-3GYZ>].

²⁹³ Human Rights Council Res. 26/9, U.N. Doc. A/HRC/RES/26/9 (July 14, 2014).

²⁹⁴ DSA, *supra* note 3. For an overview of the legislation, see *A Guide to the Digital Services Act, the EU's New Law to Rein in Big Tech*, ALGORITHM WATCH (Sept. 21, 2022), <https://algorithmwatch.org/en/dsa-explained> [<https://perma.cc/N7UV-AYNL>].

²⁹⁵ *A Guide to the Digital Services Act*, *supra* note 294; Aleksandra Kuczerawy, *Remedying Overremoval: The Three-Tiered Approach of the DSA*, VERFASSUNGSBLOG (Nov. 3, 2022), <https://verfassungsblog.de/remedying-overremoval> [<https://perma.cc/2C5M-P86Y>].

noted the relationship between the out-of-court dispute settlement provision and the Oversight Board.²⁹⁶

National jurisdictions are also increasingly active in this regulatory space. In 2019, for example, the United Kingdom released the Online Harms White Paper, which envisions legislation regulating online platforms, including through the creation of a new duty of care for social media platforms.²⁹⁷ The French Corporate Duty of Vigilance Law, adopted two years earlier, requires large French companies, or companies with large French subsidiaries, to publish a “vigilance plan” that “establish[es] effective measures to identify risks and prevent severe impacts on human rights.”²⁹⁸

For both supranational and national legislation, Oversight Board jurisprudence may provide guidance with respect to due diligence, how platforms distinguish between lawful and unlawful content, and the potential systematic effects of human rights harms created or facilitated by social media companies. In addition, national courts may look to OB decisions when interpreting such legislation or when reviewing the human rights impact of social media platforms in other areas of law.

III. RISKS AND CHALLENGES

We have thus far portrayed the Oversight Board in mostly positive terms. In particular, we have discussed the Board’s potential to enhance Meta’s accountability and to develop a case-specific jurisprudence that explains how international human rights law developed for states applies to private social media companies. In fact, the OB’s human rights future is rapidly becoming its human rights present. When we began the research for this Article, many of the proposals that we discussed seemed far off prospects. Yet the Board’s first annual report reveals that the OB is

²⁹⁶ See, e.g., Daniel Holznagel, *A Self-Regulatory Race to the Bottom Through Art. 21 Digital Services Act: How the DSA Will Introduce Competition for the Meta Oversight Board (and the German FSM) and Why We Should Be Worried About This*, VERFASSUNGSBLOG (Mar. 16, 2022), <https://verfassungsblog.de/a-self-regulatory-race-to-the-bottom-through-art-18-digital-services-act> [<https://perma.cc/VU75-4Z3U>]; Pietro Ortolani, *If You Build It, They Will Come: The DSA’s “Procedure Before Substance” Approach*, VERFASSUNGSBLOG (Nov. 7, 2022), <https://verfassungsblog.de/dsa-build-it> [<https://perma.cc/VR6F-CRVC>]; Wong & Floridi, *supra* note 3.

²⁹⁷ SEC’Y OF STATE FOR DIGIT., CULTURE, MEDIA & SPORT & SEC’Y OF STATE FOR THE HOME DEP’T, ONLINE HARMS WHITE PAPER (2019), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf [<https://perma.cc/ZL6T-BZZ4>].

²⁹⁸ *France’s Duty of Vigilance Law*, BUS. & HUM. RTS. RES. CTR., <https://www.business-humanrights.org/en/big-issues/corporate-legal-accountability/frances-duty-of-vigilance-law> [<https://perma.cc/4ZSW-DVYZ>].

2023]

META OVERSIGHT BOARD

2289

already implementing a number of these strategies as well as embracing human rights as its normative polestar.²⁹⁹

Although we are sanguine about the Oversight Board's trajectory, it is also important to acknowledge the risks and challenges that the institution is facing or may soon confront. This Part considers four of these issues: (1) pushback and backlash; (2) whitewashing; (3) indeterminacy and legitimacy; and (4) negative spillover. We describe these risks and challenges below, drawing on the experiences of international human rights tribunals and offering a few preliminary suggestions for mitigating them.

A. Pushback and Backlash

The early fears of some commentators that Meta would simply ignore costly or inconvenient Oversight Board decisions and recommendations have not come to pass. The company has complied with all OB rulings ordering it to restore or take down content and has implemented many of the Board's policy recommendations.³⁰⁰ As the Board's 2021 Annual Report notes, Meta has made significant changes in response to these recommendations, including informing users about automated review, disclosing information about government requests, and translating its rules into additional languages.³⁰¹

Meta and the OB are also having constructive interactions about the Board's future and its sometimes far-reaching policy guidance. The Annual Report asserts that in 2021, "[t]he Board got better at making recommendations and Meta improved at responding to them."³⁰² This was so even as the company complained about time and resources needed for implementation and argued that the "size and scope of the Board's recommendations go beyond the policy guidance that we first anticipated when we set up the board."³⁰³ Equally as important, Meta has agreed to

²⁹⁹ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 43 ("We believe that placing human rights at the heart of our work, and pushing Meta to respect them, will make Facebook and Instagram better places for users.")

³⁰⁰ *Id.* at 8 (reporting that with respect to the time period between October 2020 and December 2021, "the company has fully met its commitments on case decisions, has agreed to implement more than half of the Board's policy recommendations, and is increasingly answering the Board's questions"); *see also Oversight Board Response to Meta's Announcement on Reforming Its Penalty System*, OVERSIGHT BD. (Feb. 23, 2023), <https://oversightboard.com/news/507876928181835-oversight-board-response-to-meta-s-announcement-on-reforming-its-penalty-system> [<https://perma.cc/C5TH-LSVM>].

³⁰¹ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 5.

³⁰² *Id.* at 57.

³⁰³ *Id.* at 58.

provide ongoing financial support for the Board, recently pledging \$150 million to the Oversight Board Trust's initial \$130 million endowment to ensure that the OB continues beyond its six-year startup period.³⁰⁴

Yet pushback and backlash remain real concerns for the Oversight Board, as they do for international human rights tribunals.³⁰⁵ As we have explained, the company can modify the Values and Community Standards that the OB applies,³⁰⁶ although it has far less power to change the OB's jurisdiction or access rules and no control over the content of international human rights law or who is appointed to the Board. Meta can also signal (and, some would argue, is already signaling) its displeasure in less formal ways, such as by "slow walking" compliance with decisions, declining to implement policy recommendations, misreading the Board's policy decisions,³⁰⁷ and resisting its requests for information.³⁰⁸

The most notable example of pushback that the Board has encountered thus far involved Meta's decision to withdraw its request for guidance on content moderation related to the Ukraine-Russia war.³⁰⁹ The OB expressed its "disappoint[ment]"³¹⁰ in the decision—which purportedly stemmed from concerns for the safety of Meta's employees in Russia³¹¹—but did not otherwise challenge it publicly.³¹² Commentators were quick to criticize the Board's response. According to one, to ask the Board "for its advice—advice that would not even be binding on the company—and then decide belatedly that such advice might be dangerous calls into question the point of the entire

³⁰⁴ *Securing Ongoing Funding for the Oversight Board*, OVERSIGHT BD. (July 22, 2022), <https://oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board> [<https://perma.cc/8392-EUR4>].

³⁰⁵ See Madsen, Cebulak & Wiebusch, *supra* note 101, at 203.

³⁰⁶ See *supra* text accompanying note 166.

³⁰⁷ O'Kane, *supra* note 10, at 195–97 (discussing several decisions in which Meta has "mistranslate[d]" the Board's decisions, giving them a more limited reading than appears to have been intended).

³⁰⁸ See Arun, *supra* note 12, at 262 (arguing that Meta's resistance to providing information undermines the Board's legitimacy).

³⁰⁹ Casey Newton, *How Facebook Undercut the Oversight Board*, VERGE (May 12, 2022, 9:30 AM), <https://www.theverge.com/23068243/facebook-meta-oversight-board-putin-russia-ukraine-decision> [<https://perma.cc/VW35-GUDW>].

³¹⁰ Sara Fischer, *Facebook Withdraws Guidance Request for Ukraine War Content Policies*, AXIOS (May 11, 2022), <https://www.axios.com/2022/05/11/meta-facebook-oversight-board-ukraine-war> [<https://perma.cc/K4TB-MPP3>].

³¹¹ Newton, *supra* note 309 ("After asking for the opinion, the company's legal and security teams became concerned that anything the board said might somehow be used against employees or their families in Russia, either now or in the future.").

³¹² Steven Levy, *Inside Meta's Oversight Board: 2 Years of Pushing Limits*, WIRED (Nov. 8, 2022, 6:00 AM), <https://www.wired.com/story/inside-metas-oversight-board-two-years-of-pushing-limits> [<https://perma.cc/AHA6-29QJ>].

enterprise.”³¹³ Another noted that the incident “sets a precedent for Meta to continue withdrawing requests in the future.”³¹⁴

Two months after the withdrawal, however, the Board accepted a user appeal of Meta’s removal of a post comparing the Russian army in Ukraine to Nazis and quoting a poem that calls for the killing of fascists.³¹⁵ It overturned the removal and recommended that the company revise its policies to consider unlawful military intervention as relevant context.³¹⁶

Given the other areas in which the OB is incrementally “push[ing] Meta in the right direction” on human rights issues,³¹⁷ acquiescing in the company’s withdrawal of the Ukraine request while later ruling on a user appeal raising similar issues was strategically astute. The Board is likely to be more effective in enhancing Meta’s accountability to users and generating new norms if it proceeds incrementally, choosing its battles carefully and promoting its achievements to a wider audience. Such an approach may also, as we now explain, help to convince skeptics who believe that “the Oversight Board’s only role is to handle the easy questions.”³¹⁸

B. *Whitewashing*

Meta’s creation of an independent review mechanism has generated a wide range of responses. Although some human rights experts, civil society groups, and scholars have supported the initiative, participated in consultations leading to the creation of the Board, or submitted third-party interventions, others have been skeptical of or even hostile to the company’s and the Board’s embrace of human rights. A significant concern is that Meta is pledging fealty to international human rights law as a public relations fig leaf to obscure the harms of its content moderation policies and undermine prospects for meaningful regulation or reform.

According to proponents of this whitewashing claim, platforms receive “undeserved legitimacy dividends . . . [by] wrap[ping] themselves in the language of [human rights] even as what is required by that body

³¹³ Newton, *supra* note 309.

³¹⁴ Fischer, *supra* note 310.

³¹⁵ *Oversight Board Announces New Cases and Review of Meta’s COVID-19 Misinformation Policies*, OVERSIGHT BD. (July 26, 2022), <https://oversightboard.com/news/385467560358270-oversight-board-announces-new-cases-and-review-of-meta-s-covid-19-misinformation-policies> [<https://perma.cc/9M6A-GWBV>].

³¹⁶ Russian Poem, No. 2022-008-FB, 2022 Meta OB LEXIS 7, at *32–34 (Nov. 16, 2022).

³¹⁷ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 5.

³¹⁸ Newton, *supra* note 309.

of norms remains indeterminate and contested.”³¹⁹ They may also “try to co-opt the vocabulary of human rights to legitimize minor reforms at the expense of undertaking more structural or systemic changes to their moderation processes.”³²⁰ For example, the OB Charter refers not to international human rights *law*, but to human rights *norms*, perhaps indicating a reluctance by Meta to fully commit itself to legal constraints.³²¹ Lastly, skeptics charge that corporations can manipulate the rules and doctrines of human rights law—such as the legality, legitimate aim, and proportionality requirements of ICCPR Article 19—to serve their own financial and businesses interests.³²²

These are serious concerns, and Meta’s first Human Rights Report, published in July 2022, does nothing to alleviate them.³²³ According to numerous civil society critics, this self-congratulatory document “whitewash[es] a history of fomenting hatred, violence, and extremism across the globe” and “present[s] the rosier image of the company possible.”³²⁴ The most glaring omission is Meta’s failure to consider the ways in which its “surveillance advertising business model” and algorithms amplify human rights harms such as ethnic violence, hate speech, and discrimination, especially in Global South countries where its platforms are prevalent.³²⁵

The eighty-three-page report discusses the Oversight Board on only two pages (on Providing Access to Remedy). However, even in that short space, Meta both overclaims the Board’s achievements and shortchanges the contributions of other actors, extolling the OB as “defining what a human rights-based approach to content means in practice, to a level of detail, practical application and intellectual rigor, that did not previously exist.”³²⁶ Notably, Meta’s appraisal contrasts with the Board’s more

³¹⁹ douek, *supra* note 266, at 37.

³²⁰ Sander, *supra* note 35, at 1005.

³²¹ OVERSIGHT BOARD CHARTER, *supra* note 16, art. 2, § 2.

³²² See Aswad, *To Protect Freedom of Expression*, *supra* note 35; Brenda Dannecker, Neutral Governance 3 (unpublished manuscript) (on file with authors).

³²³ META, *supra* note 73.

³²⁴ Mack DeGeurin & Dell Cameron, *Experts Blast Meta’s First-Ever Human Rights Report as ‘Corporate Propaganda,’* GIZMODO (July 21, 2022), <https://gizmodo.com/facebook-meta-first-human-rights-report-myanmar-india-1849197485> [<https://perma.cc/7CZ8-VL7F>].

³²⁵ Alia Al Ghussain, *Meta’s Human Rights Report Ignores the Real Threat the Company Poses to Human Rights Worldwide*, AMNESTY INT’L (July 22, 2022), <https://www.amnesty.org/en/latest/campaigns/2022/07/metahumanrightsreport-ignores-the-real-threat-the-company-poses-to-human-rights-worldwide> [<https://perma.cc/S6E5-QY3W>]; see also Billy Perrigo, *Facebook Accused of ‘Whitewashing’ Long-Awaited Human Rights Report on India*, TIME (July 15, 2022, 8:57 AM), <https://time.com/6197154/facebook-india-human-rights> [<https://perma.cc/HQS8-QJHQ>].

³²⁶ META, *supra* note 73, at 74.

measured self-assessment in its Annual Report released a month earlier.³²⁷

We have several responses to the whitewashing critique. As an initial matter, it is important to reiterate that the Oversight Board is structurally independent from Meta. It is one thing to argue that the OB should, for example, be more stringent in reviewing the company's content moderation decisions or more forceful in its policy recommendations. It is quite another to dismiss wholesale an autonomous body that, at the very least, is nudging the platform toward greater accountability and transparency. In addition, the Board's design is remarkably free of the kinds of structural vulnerabilities that can facilitate corporate capture.³²⁸

More substantively, the Board's actions to date have narrowed Meta's ability to invoke generic references to human rights to shield itself from external scrutiny. As we have shown, OB decisions and recommendations have clarified, often with granularity, how multipart standards such as ICCPR Article 19 apply to different areas of content moderation. The Board has repeatedly insisted that Meta refrain from actions that are arbitrary, opaque, or discriminatory. In addition, as previously noted, the company has complied with all OB decisions and many policy recommendations. These are significant accomplishments for an independent monitoring body that is just a few years old. They also comprise a more substantial record of achievement than most international human rights tribunals could have claimed at similarly early points in their evolution.

Concerns about business enterprises strategically using the language of corporate social responsibility or human rights to boost their brand, improve their public image, or divert attention from the harms they cause are as old as corporate philanthropy itself.³²⁹ Yet if the Board continues along its current trajectory, and if Meta continues to comply with most of its decisions and policy recommendations—both admittedly uncertain prospects—credible charges of whitewashing are likely to diminish over time.

³²⁷ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 47 (“[U]sing international human rights standards to interpret the human rights responsibilities of a social media company is a new and challenging exercise.”).

³²⁸ See Galit A. Sarfaty & Raphaël Deberdt, *Supply Chain Governance at a Distance*, LAW & SOC. INQUIRY 16–18 (May 3, 2023), <https://doi.org/10.1017/lis.2023.17> [<https://perma.cc/SK83-Y5BQ>] (discussing how due diligence requirements regarding minerals in supply chains became tools of corporate legitimation due to the companies' role in setting, interpreting, and assessing their compliance with the relevant standards).

³²⁹ Tracey Grose, *Evolution of Corporate Philanthropy over Time*, MEDIUM (Mar. 26, 2019), <https://medium.com/@TraceyGroseSF/evolution-of-corporate-philanthropy-over-time-4a7f7bc97640> [<https://perma.cc/EGH6-FTRG>].

C. *Indeterminacy and Legitimacy*

Another concern relates to the indeterminacy of human rights law and the legitimacy of the Oversight Board's attempts to provide greater clarity. On the one hand, skeptics claim that indeterminacy makes it impossible for the OB to impose meaningful constraints on Meta. Yet they also fear that if the Board clears up these ambiguities, it will become "the ultimate arbiter of free speech norms around the world."³³⁰ Such a role would, in their view, be illegitimate: the OB is neither a national court nor an international tribunal but a private oversight body comprised of unaccountable, self-appointed experts whose decisions cannot be challenged or corrected by states, civil society, or Meta users.³³¹ This set of challenges thus asserts, somewhat paradoxically, that the Oversight Board will either have too little influence—or too much. We believe both dimensions of this critique are overstated.

Consider first indeterminacy. Skeptics argue that human rights law is too ambiguous and internally inconsistent to guide Meta's content moderation decisions or the Board's review of those determinations. Proponents of this view claim that there is "no accepted comprehensive set of global standards that can operate as the default rules for content moderation."³³²

The texts of human rights treaties do lack precision and granularity, at least with regard to provisions—such as freedom of expression, privacy, and nondiscrimination—that are most relevant to content

³³⁰ douek, *supra* note 9, at 7.

³³¹ Dvoskin critiques the Board's reliance on international human rights law as enabling it to present its decisions as apolitical and incontestable. See generally Brenda Dvoskin, *Expert Governance of Online Speech*, 64 HARV. INT'L L.J. 85 (2023). Claims that human rights are hegemonic have a long history. See, e.g., David Kennedy, *The International Human Rights Movement: Part of the Problem?*, 15 HARV. HUM. RTS. J. 101, 108 (2002). Whatever the potential merits of these claims in the abstract, they lack force at the level of institutions, which do not seek to establish incontestable truths but rather to provide remedies, promote accountability, and offer a way for claimants to articulate legal and political demands. See, e.g., Michael Goodhart, *Human Rights and the Politics of Contestation*, in HUMAN RIGHTS AT THE CROSSROADS 31 (Mark Goodale ed., 2013). Analogizing the Board to international human rights tribunals helps to illuminate whether, and, if so, how the OB can serve these functions.

³³² Brenda Dvoskin, *Why International Human Rights Law Cannot Replace Content Moderation*, MEDIUM (Oct. 8, 2019), <https://medium.com/berkman-klein-center/why-international-human-rights-law-cannot-replace-content-moderation-d3fc8dd4344c> [https://perma.cc/U8FC-ASE2]. But see Evelyn Aswad & David Kaye, *Convergence & Conflict: Reflections on Global and Regional Human Rights Standards on Hate Speech*, 20 NW. J. HUM. RTS. 165, 168 (2022) ("U.N. treaty bodies, along with expert opinions offered within the U.N. system, have developed an increasingly consistent set of rules governing the appropriate boundaries for hate speech laws.").

moderation.³³³ Yet this is a feature of individual rights protections in general. A modest amount of indeterminacy allows international tribunals and domestic courts to clarify and update rights in light of contemporary challenges and normative developments. It also allows decisionmakers and advocates to adopt interpretations of rights that are suitable for local contexts.

This is especially true with respect to freedom of expression. A set of global rules that purported to draw hard and fixed lines between permitted and prohibited speech and apply them across all national contexts would be both normatively problematic and practically unworkable given the diverse views regarding the benefits and harms of speech that often exist both within a single society and across countries.³³⁴ Frameworks that channel restrictions through multipart standards, such as those in ICCPR Article 19, strike a different balance: they facilitate challenges to abusive and arbitrary speech restrictions while accommodating a modicum of political and cultural variation and preferences of different polities.³³⁵

Critics of indeterminacy also overlook the fact that treaty texts are only the starting point of what international human rights law requires. Just as the scope of domestic speech protections cannot be determined by reviewing constitutions without also examining the caselaw of the courts applying them, so too treaties protecting free expression cannot be understood without consulting the interpretations of human rights tribunals and experts. A difference in the international context is that a multiplicity of institutions engages in this interpretive exercise. The relationship among these bodies is pluralistic rather than hierarchical;³³⁶ this, too, is a feature, not a bug. An extensive literature on fragmentation and harmonization explains how interactions among these institutions over time lead to greater particularity and clarity to shared legal texts,

³³³ See International Covenant on Civil and Political Rights, *supra* note 83, arts. 2, 17, 19.

³³⁴ SVEND-ERIK SKAANING & SUTHAN KRISHNARAJAN, WHO CARES ABOUT FREE SPEECH? FINDINGS FROM A GLOBAL SURVEY OF SUPPORT FOR FREE SPEECH 8–13 (2021), https://futurefreespeech.com/wp-content/uploads/2021/06/Report_Who-cares-about-free-speech_21052021.pdf [<https://perma.cc/CX8W-FQH7>].

³³⁵ See, e.g., Aswad & Kaye, *supra* note 332, at 210–11 (explaining that the three-part test in ICCPR Article 19 “do[es] not force the homogenization of all speech laws” but “in fact require[s] a robust examination of local conditions and context in order to assess the validity of speech restrictions”).

³³⁶ See, e.g., Nico Krisch, *The Pluralism of Global Administrative Law*, 17 EUR. J. INT’L L. 247, 248 (2006) (arguing that pluralism is a “normatively adequate response” to “fundamental and durable contestation over the right constituency of global governance”).

even as some divergences remain at the margins.³³⁷ We expect that OB decisions and recommendations will contribute to this interpretive enterprise, providing much-needed specificity to how human rights law applies to content moderation decisions across a range of factual contexts.

Yet skeptics also raise the opposite critique, challenging the Oversight Board's efforts to ameliorate the indeterminacy of human rights standards. According to this view, the Board's top-down, expert-driven attempts to balance the benefits and harms of speech in different contexts are suspect.³³⁸ At a minimum, they are not entitled to greater respect than any other reasonable approach to balancing.

In our view, the legitimacy of the Board's normative output depends on the audience. Meta users agree to be governed by the company's Community Standards and Values and by the OB's review of those standards—including under international human rights law—when they create Facebook or Instagram accounts.³³⁹ They also agree that OB decisions are binding on the company. The large and growing number of appeals to the OB suggests that many users see value in challenging Meta's content moderation decisions. To be sure, the decision to join the platform does not preclude users from objecting to the Board's legitimacy, any more than the formal consent of states resolves all legitimacy concerns relating to international institutions.³⁴⁰ But it does mean that such challenges should be evaluated differently than legitimacy objections raised by other actors affected by Meta's system of content moderation—in particular, non-users of Facebook and Instagram.

In assessing the Board's legitimacy for external audiences, it is also important to underscore that the Board's interpretations of international human rights law are only persuasive authority.³⁴¹ Its decisions do not

³³⁷ See generally BEYOND FRAGMENTATION: CROSS-FERTILIZATION, COOPERATION AND COMPETITION AMONG INTERNATIONAL COURTS AND TRIBUNALS (Chiara Giorgetti & Mark Pollack eds., 2022); Laurence R. Helfer, *Forum Shopping for Human Rights*, 148 U. PA. L. REV. 285 (1999).

³³⁸ See, e.g., Bloch-Wehba, *supra* note 12, at 93 (characterizing the OB as a “simulacrum of due process, unregulated by law or the Constitution, and therefore, unaccountable to the democratic process”); Dannecker, *supra* note 322, at 1 (arguing that the OB “may legitimize experts’ power at the expense of pluralistic politics”).

³³⁹ *Terms of Service*, FACEBOOK, <https://www.facebook.com/terms.php> [<https://perma.cc/V88Q-JDNX>]; *Terms of Use*, INSTAGRAM, https://help.instagram.com/581066165581870/?helpref=uf_share [<https://perma.cc/MUQ3-Y5V3>].

³⁴⁰ See, e.g., Daniel Bodansky, *The Legitimacy of International Governance: A Coming Challenge for International Environmental Law?*, 93 AM. J. INT’L L. 596, 597 (1999).

³⁴¹ See, e.g., Fertmann and Tiedeke, *supra* note 285, at 23 (2022) (“[T]he Oversight Board carefully crafts itself as a persuasive international human rights interpreting institution.”); Klonick, *supra* note 7, at 2495 (“Courts are unlikely to accept the Board’s decision as precedential or give it comity, but they might consider it persuasive.”).

bind any other institution or actor. Rather, the OB's normative influence will depend on a range of factors that scholars of the legitimacy of international tribunals have identified, such as whether it follows fair and transparent processes, whether it considers the views of a diverse range of affected actors, whether its decisions are well reasoned and persuasive, and whether it continues to maintain its independence from Meta.³⁴²

D. *Negative Spillover*

Although we are ultimately unpersuaded by the indeterminacy and legitimacy critiques of the Oversight Board's human rights future, we acknowledge a related concern. The OB is a private monitoring body tasked with generating a robust body of public law norms on a relatively narrow and underdeveloped topic. It has extensive resources to generate these norms quickly and disseminate them widely, and it is actively working to increase its output.³⁴³ Although we previously championed a vision of the Board's salutary contributions to human rights, there are also risks of negative spillover from a body with these characteristics. In particular, the OB's decisions could generate norms that are appropriate when applied to Meta but are mischaracterized or abused by governments or other social media platforms in ways that will be difficult for international human rights tribunals to rectify.³⁴⁴

A useful example is the line of OB decisions holding that Meta can remove offensive expressions such as a racial slur, an image of blackface, and a rat cartoon depicting an ethnic group. In each instance, the Board recognized that Meta's Community Standard for hate speech "prohibits specific forms of discriminatory expression" that "would raise concerns

³⁴² See HARLAN GRANT COHEN, ANDREAS FOLLESDAL, NIENKE GROSSMAN & GEIR ULFSTEIN, *Legitimacy and International Courts—a Framework*, in LEGITIMACY AND INTERNATIONAL COURTS 1 (Nienke Grossman, Harlan Grant Cohen, Andreas Follesdal & Geir Ulfstein eds., 2018).

³⁴³ The Board recently announced changes to its Charter and Bylaws allowing expedited and summary decisions that will enable it "to review more cases and to do so faster than before." *Oversight Board Announces Plans to Review More Cases, and Appoints a New Board Member*, OVERSIGHT BD. (Feb. 14, 2023), <https://oversightboard.com/news/943702317007222-oversight-board-announces-plans-to-review-more-cases-and-appoints-a-new-board-member> [<https://perma.cc/4M4S-GJLT>].

³⁴⁴ See Aziz Z. Huq, *International Institutions and Platform-Mediated Misinformation*, 23 CHI. J. INT'L L. 116, 123 (2022) ("[O]nce content moderation is hitched to human rights, there is a new, potentially powerful incentive to capture the bodies that generate such law."); David Kaye, *Platform Rules and Public Norms* (U.C. Irvine Sch. of L., Legal Studies Research Paper Series No. 2022-26, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4180964 [<https://perma.cc/9GUB-GNUK>] (exploring how private companies justifying their actions in public law terms might influence public law and policies).

if imposed by a government at a broader level.”³⁴⁵ Nevertheless, the OB concluded that the company “can regulate such expression[by] demonstrating the necessity and proportionality of the action.”³⁴⁶ The Board recently articulated this principle in more general terms: “Meta can apply less strict standards for removing content from its platform than those which apply to states imposing criminal or civil penalties.”³⁴⁷

At one level, these statements simply reflect the widely shared view that public and private actors are not similarly situated when it comes to restricting freedom of expression. For one, private actors have their own speech interests and thus “can limit, curate, and privilege content and viewpoints in a way that a state should not.”³⁴⁸ For another, a social media platform that removes an individual user’s post after evaluating its impact in a particular context causes substantially less harm (at least in most instances) than a statute or executive decree that punishes broad categories of disfavored speech or speakers.

Yet states and businesses could invoke the Board pronouncements quoted above in troubling ways. Consider first responses by states. At the most general level, governments may chafe at a legal standard that gives greater leeway to one of the world’s largest and richest companies. Officials might respond to this imbalance by invoking Board decisions to support a weakening of public speech restrictions.³⁴⁹ Such a risk is especially acute given that government regulation of internet platforms often blurs the line between public and private. Even more troublingly, a state might misconstrue the Board’s focus on criminal or civil penalties as tacitly condoning political pressure to remove disfavored speakers or content from online forums, so long as laws on the books do not threaten jail time, fines, or other legal consequences for violators.

Turning to private actors, social media companies that have little interest in fact-specific determinations about whether to remove content may cite the OB’s more deferential review of removal decisions by Meta—which is at least nominally concerned with making such fine-grained distinctions—to justify removing expression from their platforms without applying the legality, legitimacy, and proportionality standard. Firms in other industries may also invoke Board decisions as support for speech restrictions, such as limits on employee tweets or letters to the

³⁴⁵ *Knin Cartoon*, No. 2022-001-FB, 2022 Meta OB LEXIS 3, at *34–35 (June 17, 2022); *South African Slurs*, No. 2021-011-FB, 2021 WL 4822229, at *10 (Sept. 28, 2021).

³⁴⁶ *Knin Cartoon*, 2022 Meta OB LEXIS 3, at *35.

³⁴⁷ *Id.* at *37.

³⁴⁸ *Id.*, *supra* note 266, at 61.

³⁴⁹ This risk exists both for existing restrictions on freedom of expression (such as ICCPR Article 19) and for future international lawmaking initiatives (such as the draft treaty on business and human rights).

2023]

META OVERSIGHT BOARD

2299

editor, on matters of public interest that relate to a company's business. A different type of negative spillover could occur if corporations cite the Board's pronouncements to seek a watering down of the human rights standards in the UNGPs.

None of this parade of horrors is inevitable. To reduce the risk of these outcomes, the Oversight Board needs, at a minimum, to be aware that its decisions are likely to be scrutinized and invoked by public and private actors other than Meta.³⁵⁰ Armed with that knowledge, the OB can provide concrete guidance as to the scope of the company's discretion to remove content in ways that may limit the ability of other actors to mischaracterize or abuse the legal principles it develops.

CONCLUSION

The Meta Oversight Board will not solve the many challenges of content moderation at scale, let alone the public law challenges of private companies governing the speech of billions of people around the world. Addressing these issues requires regulation not only of downstream outputs but also of upstream choices regarding the design of institutions and technology.³⁵¹ It also requires a good deal of patience. If the experience of international tribunals teaches us anything, it is that human rights accountability and norm development are incremental processes that require a plurality of approaches.

We argue that the Board has the potential to serve as an important check on Meta and to significantly advance the promotion and protection of rights online. We make this case by analogizing the OB to international human rights tribunals and exploring what the tribunals' trajectories suggest about how the Board can hold Meta accountable and contribute to ongoing efforts to regulate social media platforms. In doing so, this Article provides the first examination of the human rights origins of the OB, the strategies it is using to pressure Meta to improve content moderation, and how the Board is developing human rights norms applicable to private social media companies. We complement this analysis with a range of recommendations about how the Board can be even more effective and have a more lasting impact.

³⁵⁰ The Board's first Annual Report highlights several "national and international institutions [that] have started to analyze the Board's work and decisions." OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 44. None of the examples cited involve misapplications of OB standards.

³⁵¹ douek, *supra* note 20, at 532.

This experiment in governing the “governors”³⁵² has already yielded significant benefits. In just a few years, the Oversight Board has explored how human rights apply differently to corporations than governments, forced Meta to be more transparent with users in concrete ways, pushed the company to consider different types of contexts when moderating content, investigated how Meta responds to government requests to remove user posts, prompted revision of Meta’s policy on penalties, and weighed in on Meta’s cross-check program. To some observers, these may seem small steps when compared to the urgent need for greater user protections in the digital realm. We share this concern. It is for this reason that we envision the Board as only one of many public and private efforts—at the national, regional, and global levels—to push Meta and other social media platforms toward greater respect for human rights. However, when compared to the incremental and hard-fought accomplishments of international tribunals, the Board’s achievements are already impressive.

While we urge the Board to continue along this trajectory, the OB’s early successes also give us pause. A de facto human rights tribunal created by an exceedingly powerful technology company and equipped with impressive resources raises two related concerns. First, the Board is the creation of a social media giant, whose motto was at one point “move fast and break things.”³⁵³ One might argue that the international human rights system moves too slowly, but there are also risks in moving too quickly. Although speed helps the Board to keep up with technological developments, it can also lead to blind spots, missed opportunities, and inadequately reasoned decisions. In addition, given the Board’s extensive resources and the pressure it faces to demonstrate results, its efforts may quickly outstrip the normative output and monitoring efforts of established international human rights bodies.

Second, tech companies have a poor record of trying to “solve” human rights or humanitarian problems. These efforts tend to focus on quick-fix technological solutions—shiny new things that suck all the air (and funding) out of the room—rather than long-term investments in infrastructure, the rule of law, or broader social and political change.³⁵⁴ There is a risk that the Board will become a similar “technology of law”

³⁵² Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018).

³⁵³ Isobel Asher Hamilton, *Mark Zuckerberg’s New Values for Meta Show He Still Hasn’t Truly Let Go of ‘Move Fast and Break Things,’* BUS. INSIDER (Feb. 16, 2022, 11:05 AM), <https://www.businessinsider.com/meta-mark-zuckerberg-new-values-move-fast-and-break-things-2022-2> [<https://perma.cc/JT8C-5PN8>].

³⁵⁴ See generally EVGENY MOROZOV, TO SAVE EVERYTHING, CLICK HERE: THE FOLLY OF TECHNOLOGICAL SOLUTIONISM (2013).

2023]

META OVERSIGHT BOARD

2301

that crowds out other potential solutions. Some of the OB's rhetoric suggests this possibility. The Board describes itself as "a source of best practice for self-regulation," and it promises to "share [its] expertise about self-regulatory solutions across tech policy conversations, as well as contribute to discussions about content moderation innovation with potential tech industry partners."³⁵⁵ These statements could be read as a vision for a dominant institution that overshadows and diverts resources from other efforts to hold states and private companies accountable for rights violations, both online and off, or preempts efforts to address more directly the underlying structures of our attention economy and their consequences. Avoiding these pitfalls requires the Board to remain aware of its own limitations.

Notwithstanding these cautionary words, we believe the Board is already making noteworthy contributions to advancing human rights. As it continues to provide Meta with interpretive guidance and recommendations, the OB will gradually shape the company's content moderation decisions and policies as well as international human rights law writ large. The Board may also assist national and regional lawmakers and courts in regulating social media companies and particularizing their due diligence obligations. And Meta's engagement with the OB may affect the company's culture, helping to inculcate the values of transparency, information sharing, and fair process.

The Oversight Board's most lasting impact, however, will be its contribution to the accountability efforts of other actors. The experiences of international human rights tribunals reveal that long-term change requires deep and genuine engagement with civil society organizations, advocacy groups, and other supporters. The Board's decisions will thus have the greatest impact if they are deployed by these actors to press for meaningful change by states and companies alike.

³⁵⁵ OVERSIGHT BOARD ANNUAL REPORT, *supra* note 18, at 68.