

**FACEBOOK’S “OVERSIGHT BOARD:” MOVE FAST WITH STABLE
INFRASTRUCTURE AND HUMILITY**

*Evelyn Douek**

Facebook’s proposed Oversight Board is one of the most ambitious constitution-making projects of the modern era. With pre-existing governance of tech platforms delegitimized in the ongoing “techlash,” this represents a pivotal moment when new constitutional forms can emerge that will shape the future of online discourse. For all the potential of the Facebook Oversight Board (FOB), there are many things it cannot be. It will not hear a sufficient proportion of Facebook’s content moderation cases to be a meaningful response to calls for greater due process in individual decisions. Nor will it be able to become a font of globally accepted speech norms for the worldwide platform. The true value that the Board can bring to Facebook’s content moderation ecosystem lies between these two extremes of individual error correction and the settlement of globally applicable speech rules. The institutional offering of the Board should focus on two primary, but more modest, functions. First, it can help highlight weaknesses in the policy formation process at Facebook, removing blockages (such as blind spots and inertia) in the “legislative process” leading to the formulation of its Community Standards. Second, by providing an independent forum for the discussion of disputed content moderation decisions, the Board can be an important forum for the public reasoning necessary for persons in a pluralistic community to come to accept the rules that govern them, even if they disagree with the substance of those rules. Understanding the institutional role of the Board in these terms provides useful insights into the institutional design that will best help it achieve these goals.

* Doctoral candidate, Harvard Law School. B. Com./LL.B. UNSW (Hons. I). Thanks are owed to Martha Minow, Jack Goldsmith, Cass Sunstein, Noah Feldman, Thomas Kadri, Elettra Bietti, Rafe Andrews and the editors of N.C. JOLT. I can’t fob my errors off onto anyone else though; they remain my own.

I.	INTRODUCTION.....	2
II.	THE FOB’S PURPOSE	9
	<i>A. Facebook’s Approach to the Impossible Task of Content Moderation.....</i>	<i>10</i>
	<i>B. The Benefits of Voluntary Restraints</i>	<i>15</i>
	1. <i>Bestowing Legitimacy and Reassuring Users.....</i>	<i>18</i>
	2. <i>Staving Off or Guiding Regulation</i>	<i>21</i>
	3. <i>Outsourcing Controversy.....</i>	<i>24</i>
	4. <i>Enforcing Existing Standards</i>	<i>26</i>
III.	THE FOB BLUEPRINT	28
	<i>A. Membership</i>	<i>29</i>
	<i>B. Power of Review</i>	<i>34</i>
	<i>C. Subject-Matter Jurisdiction</i>	<i>37</i>
IV.	THE LEGITIMACY CONUNDRUM.....	46
	<i>A. The Limits of the FOB’s Legitimizing Power</i>	<i>47</i>
	<i>B. The Limits of a “Constitution’s” Legitimizing Power....</i>	<i>49</i>
V.	MAXIMIZING THE FOB’S POTENTIAL	53
	<i>A. Weakness as Strength.....</i>	<i>53</i>
	1. <i>The Benefits of a Judicial-Style Check on Policy-Making</i>	<i>54</i>
	2. <i>The Advantages of Weak-Form Review</i>	<i>58</i>
	3. <i>A Check on Users Too</i>	<i>64</i>
	<i>B. Public Reasoning.....</i>	<i>66</i>
	1. <i>The Purpose of Public Reasons</i>	<i>67</i>
	2. <i>The Challenge of Giving Acceptable Reasons</i>	<i>73</i>
VI.	CONCLUSION	76

I. INTRODUCTION

Mark Zuckerberg, the Founder and Chief Executive Officer of Facebook, is engaged in one of the most ambitious constitutional projects of the modern era. On November 15, 2018, he announced in a blog post that by the end of 2019, Facebook will “create a new way for people to appeal content decisions to an independent body,

whose decisions would be transparent and binding.”¹ Facebook has since released a “Draft Charter,” describing the “Oversight Board” as “a body of independent experts who will review Facebook’s most challenging content decisions – focusing on important and disputed cases.”² In an earlier interview, when he first floated the idea, Zuckerberg analogized the proposed body to a “Supreme Court.”³ Thus, it seems that Zuckerberg is intending to introduce a check and balance into the governance of his sovereign domain of “Facebookistan.”⁴

This innovation comes amidst ongoing constitutional upheaval for the Internet. The “techlash”⁵ of the past few years, started by revelations of fake news and disinformation in the 2016 U.S. election⁶ but sweeping broadly through the tech sector since then,⁷

¹ Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (Nov. 15, 2018), <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/> [https://perma.cc/C7P3-DLYT].

² DRAFT CHARTER: AN OVERSIGHT BOARD FOR CONTENT DECISIONS, FACEBOOK (Jan. 28, 2019) [hereinafter DRAFT CHARTER]. This paper reflects publicly-available details as of September 10, 2019.

³ Ezra Klein, *Mark Zuckerberg on Facebook’s hardest year, and what comes next*, VOX (Apr. 2, 2018), <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge> [https://perma.cc/7A2Z-HT6D].

⁴ See REBECCA MACKINNON, CONSENT OF THE NETWORKED: THE WORLDWIDE STRUGGLE FOR INTERNET FREEDOM 149 (2012); Anupam Chander, *Facebookistan*, 90 N.C. L. REV. 1807 (2012).

⁵ Rana Foroohar, *Year in a Word: Techlash*, FIN. TIMES (Dec. 16, 2018), <https://www.ft.com/content/76578fba-fca1-11e8-ac00-57a2a826423e> [https://perma.cc/8RJH-CV2B]; Eve Smith, *The Techlash Against Amazon, Facebook and Google—And What They Can Do*, ECONOMIST (Jan. 20, 2018), <https://www.economist.com/briefing/2018/01/20/the-techlash-against-amazon-facebook-and-google-and-what-they-can-do> [https://perma.cc/BBH6-LNTY]; *Word of the Year 2018: shortlist*, OXFORD DICTIONARIES, <https://en.oxforddictionaries.com/word-of-the-year/shortlist-2018> [https://perma.cc/G9D4-GTBV].

⁶ OFFICE OF THE DIR. OF NAT’L INTELLIGENCE, BACKGROUND TO “ASSESSING RUSSIAN ACTIVITIES AND INTENTIONS IN RECENT US ELECTIONS:” THE ANALYTIC PROCESS AND CYBER INCIDENT ATTRIBUTION (Jan. 6, 2017).

⁷ Facebook has been under fire for everything from failing to prevent political operatives from having access to private data, to having a “determining role” in the ongoing genocide in Myanmar, to “poisoning” or “breaking” democracy.

has disrupted the status quo. Comparative scholarship shows that “[n]ew constitutional forms emerge only under extraordinary historical conditions, moments when pre-existing political, economic, and social structures have been weakened [or] delegitimized.”⁸ With prior content moderation practices receiving unprecedented public scrutiny and suffering a crisis of legitimacy, this moment represents such conditions for online platforms. The changes that emerge now will have significant ramifications for the future of online discourse, and because these online platforms mediate so much of modern life, sometimes called the “modern public square,”⁹ the effects will be far reaching.

Facebook is the most globally dominant social media company, with 2.32 billion monthly active users in countries around the world using the service in over 100 different languages.¹⁰ The public discourse that occurs on Facebook is central to political and cultural life in and between countries around the world.¹¹ As tech companies everywhere are looking for solutions to the problems of content moderation, Facebook’s governance innovation could provide a model—or cautionary tale—for others. This is therefore a pivotal moment in the history of online speech governance.

Kevin Granville, *Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens*, N.Y. TIMES (May 12, 2018), <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html> [https://perma.cc/5WGE-D9RK]; REUTERS, *Myanmar: UN Blames Facebook for Spreading Hatred of Rohingya*, GUARDIAN (Mar. 12, 2018), <http://www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya> [https://perma.cc/YXH9-ZJM7]; Zeynep Tufekci, *It’s the (Democracy-Poisoning) Golden Age of Free Speech*, WIRED (Jan. 16, 2018), <https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/> [https://perma.cc/LJB3-PPLW]; Evan Osnos, *Can Mark Zuckerberg Fix Facebook Before It Breaks Democracy?*, NEW YORKER (Sept. 17, 2018), <https://www.newyorker.com/magazine/2018/09/17/can-mark-zuckerberg-fix-facebook-before-it-breaks-democracy> [https://perma.cc/YUR5-BNHN].

⁸ ALEC STONE SWEET, *GOVERNING WITH JUDGES: CONSTITUTIONAL POLITICS IN EUROPE* 38 (2000).

⁹ *Packingham v. North Carolina*, 582 U.S. ___, ___, 8 (2017).

¹⁰ FACEBOOK NEWSROOM, <https://newsroom.fb.com/company-info/> [https://perma.cc/BL5F-NR8S] (last visited Sept. 22, 2019).

¹¹ ZEYNEP TUFEKCI, *TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST* 6 (2017).

For all the potential of the Facebook Oversight Board (hereinafter, FOB), there are many things it cannot be. Politicians,¹² academics,¹³ civil society,¹⁴ and users¹⁵ have long been calling for greater transparency and due process in Facebook's application of its Community Standards and other content moderation decisions.¹⁶ However, the FOB cannot be a meaningful answer to these calls. In other contexts, an appeal or some form of "judicial review" of a decision can be a form of due process: such mechanisms give complainants an opportunity to voice their grievance, have a hearing, and receive some form of explanation for their treatment.¹⁷ Appeals processes can also be a way of ensuring the effective functioning of a bureaucratic system and rule enforcement by creating a mechanism for error correction. But the sheer volume of content moderation decisions Facebook makes every day¹⁸ means that the FOB cannot be expected to offer this kind of procedural recourse or error correction in anything but the smallest fraction of

¹² Evelyn Douek, *Transatlantic Techlash Continues as U.K. and U.S. Lawmakers Release Proposals for Regulation*, LAWFARE (Aug. 8, 2018), <https://www.lawfareblog.com/transatlantic-techlash-continues-uk-and-us-lawmakers-release-proposals-regulation> [<https://perma.cc/5CYR-LFJX>].

¹³ Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011 (2018); Hannah Bloch-Wehba, *Global Platform Governance and Private Power in the Shadow of the State*, 72 SMUL. REV. 66 (2019); Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035 (2018).

¹⁴ *An Open Letter to Mark Zuckerberg*, SANTA CLARA PRINCIPLES (Nov. 13, 2018), <https://santaclaraprinciples.org/images/scp-og.png> [<https://perma.cc/EEQ5-ETFM>]; David Kaye, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. Doc. A/HRC/38/35 (June 4, 2018).

¹⁵ See, e.g., Civil society organizations, on behalf of 143,000 people, Letter to Mark Zuckerberg (Aug. 22, 2016) https://s3.amazonaws.com/s3.sumofus.org/images/FinalLetter-MarkZuckerberg_1.pdf [<https://perma.cc/9ZJS-ST5U>].

¹⁶ See TIMOTHY GARTON ASH ET AL., GLASNOST! NINE WAYS FACEBOOK CAN MAKE ITSELF A BETTER FORUM FOR FREE SPEECH AND DEMOCRACY 18–20 (2019).

¹⁷ Yuval Eylon & Alon Harel, *The Right to Judicial Review*, 92 VA. L. REV. 991, 997 (2006).

¹⁸ See discussion *infra* at Part II(A).

these cases.¹⁹ This is compounded in Facebook’s case by the difficulty of ensuring the FOB’s decisions are effectively communicated and absorbed by the globally distributed and time-starved workforce of content moderators that make the first instance content moderation decisions.²⁰

Tellingly, due process or individual error correction are not the reasons Facebook says it is establishing the FOB. “Due process” is not mentioned in Zuckerberg’s blog post announcing the blueprint for the body or in the Draft Charter released a few months later.²¹ As for error correction, Zuckerberg says that artificial intelligence and more content moderators will help reduce mistakes but “we will never be perfect.”²² Facebook seems to conceive of the FOB as something different, and grander. In describing the function of the body, the Draft Charter explains that because decisions over what should and should not be allowed on Facebook are “too consequential for Facebook to make alone”²³ the FOB will review the “most challenging content decisions” to provide oversight and make Facebook more accountable.²⁴ It will be charged with reviewing enforcement of Facebook’s Community Standards—which “apply around the world to all types of content”²⁵—in

¹⁹ Issie Lapowsky, *Real Facebook Oversight Requires More Than a 40-Expert Board*, WIRED (Jan. 28, 2019), <https://www.wired.com/story/facebook-oversight-board-draft-charter/> [<https://perma.cc/ZYU9-F6JB>].

²⁰ See, e.g., Sarah T. Roberts, *Social Media’s Silent Filter*, ATLANTIC (Mar. 8, 2017), <https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/> [<https://perma.cc/3EEG-T28K>] (“In a matter of seconds, following pre-determined company policy, CCM workers make decisions about the appropriateness of images, video, or postings that appear on a given site . . . Increasingly, CCM work is done globally, . . . CCM workers are almost always contractors, in many cases limited in term due to their high rate of burnout.”).

²¹ Zuckerberg, *supra* note 1.

²² *Id.*

²³ Nick Clegg, *Charting a Course for an Oversight Board for Content Decisions*, FACEBOOK NEWSROOM (Jan. 28, 2019) <https://newsroom.fb.com/news/2019/01/oversight-board/> [<https://perma.cc/9A4W-XAM2>].

²⁴ DRAFT CHARTER, *supra* note 2, at 1.

²⁵ *Community Standards*, FACEBOOK, <https://www.facebook.com/communitystandards/introduction/> [<https://perma.cc/5ZXL-VTF8>] (last visited Sept. 22, 2019).

accordance with a set of indisputably worthy and sweeping values “like voice, safety, equity, dignity, equality and privacy.”²⁶ However, if the vision of the FOB is to be a magisterial font of global speech norms for the global platform, this too is mistaken. The FOB will never be the ultimate arbiter of free speech norms around the world whose pronouncements are accepted as legitimate in the way a court of final appeal might in a domestic legal system. It has neither the legitimacy nor authority necessary to fulfil this role.

The true value that the FOB can bring to Facebook’s content moderation ecosystem lies between these two extremes of individual error correction and the settlement of globally applicable speech rules. The institutional offering of the FOB should focus on two primary, but more modest, functions. First, it can help highlight weaknesses in the policy formation process at Facebook, removing blockages (such as blind spots and inertia) in the “legislative process” leading to the formulation of its Community Standards.²⁷ Second, by providing an independent forum for the discussion of disputed content moderation decisions, the FOB can be an important forum for the process of public reasoning necessary for persons in a pluralistic community to come to accept the rules that govern them, even if they disagree with the substance of those rules.²⁸ Understanding the institutional role of the FOB in these terms provides useful insights into the institutional design that will best help it achieve these goals.

This article proceeds in four parts. Part II looks at the context for the introduction of the FOB into Facebook’s content moderation ecosystem. Content moderation—determining what should and shouldn’t be allowed on Facebook’s platform—is both an impossible and indispensable part of Facebook’s business. This paradox sheds light on why Facebook would voluntarily subject itself to the constraints of an independent appeal body. Key amongst the reasons for establishing the FOB is the desire to find a way of

²⁶ DRAFT CHARTER, *supra* note 2, at 3.

²⁷ Rosalind Dixon, *The Core Case for Weak-Form Judicial Review*, 38 CARDOZO L. REV. 2193 (2017).

²⁸ JOHN RAWLS, POLITICAL LIBERALISM 212 (2005).

legitimizing the power that Facebook exercises over its users and the public sphere. It is only in understanding this context that we can begin to ask what would make the FOB's institutional design effective.

Part III sketches the current vision of the FOB based on early documents describing its institutional design. While not final, these documents reveal key design decisions, such as the intended membership of the FOB, its power to review Facebook decisions, and the scope of its jurisdiction.

Part IV turns to the limitations on the FOB and the ways in which it differs from any prior institution that make the legitimizing goal especially difficult for the FOB. Not only will it be difficult for Facebook to make credible its commitment to be bound by the FOB in cases where its decisions are damaging to the company's business interests, but the FOB also does not have a reservoir of legitimacy to draw on when making deeply contested decisions. The FOB will need to establish its legitimacy over time, and its task will be all the more challenging in these circumstances.

Part V turns to the ways in which Facebook can maximize the value the FOB can bring to its content moderation ecosystem in light of these inherent limitations. The value of a judicial-style check does not require the FOB to have ultimate say in all decisions; indeed, a "weak-form" judicial review model may be more appropriate in the dynamic environment of online speech. Nor does the FOB need to issue universally-accepted decisions; indeed, it will not be able to. But it is the process of public reasoning itself which is the *raison d'être* of the FOB and should be the focus of the body's design.

The FOB is an experiment in governance that responds to an unprecedented amount of control by a private platform over the global public sphere. It will not and should not be a copy of any institution from a radically different context. This does not mean that there is not much that can be learned from previous experience before the beta version of this updated form of platform governance is released. This paper explores those lessons.

II. THE FOB'S PURPOSE

On its face, Zuckerberg and Facebook's decision to set up the FOB might seem counterintuitive. Content moderation—platforms' practice of designing and enforcing rules for what they allow to be posted on their services—is, as Gillespie says, *the* commodity that platforms offer users.²⁹ The decision to move some of the power over this core part of Facebook's business into independent hands may seem puzzling. Public pressure over Facebook's decisions in recent years is not enough to explain the move. Zuckerberg is insulated from the need to respond to public pressure by his near absolute control of the company as founder, chief executive, chairman of the Board, and majority shareowner.³⁰ But as Facebook's Head of Policy Management has written: “[s]imply put, there are business reasons that a big social media company must pay attention to what the world thinks of its speech rules.”³¹ The goals of the FOB need to be understood in the context of how it might serve these business imperatives.

There is a much more familiar context in which holders of expansive unilateral power nevertheless employ a court system as a check and balance: rulers in authoritarian regimes. Scholars have previously drawn the comparison of Facebook as an autocracy³² due

²⁹ TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 5 (2018).

³⁰ Julia Carrie Wong, *Zuckerberg's Control of Facebook is near Absolute – Who Will Hold Him Accountable?*, GUARDIAN (Nov. 21, 2018), <http://www.theguardian.com/technology/2018/nov/21/zuckerbergs-control-of-facebook-is-near-absolute-who-will-hold-him-accountable> [<https://perma.cc/2ZBH-KPVY>].

³¹ Monika Bickert, *Defining the Boundaries of Free Speech on Social Media*, in THE FREE SPEECH CENTURY 254, 265 (Lee C. Bollinger & Geoffrey R. Stone eds., 2018).

³² David Pozen, *Authoritarian Constitutionalism in Facebookland*, KNIGHT FIRST AMEND. INST. (Oct. 30, 2018), <https://knightcolumbia.org/content/authoritarian-constitutionalism-facebookland> [<https://perma.cc/PP95-Z3RD>] (drawing on Mark Tushnet's characterization of absolutist constitutionalism as a system as one where “a single decisionmaker motivated by an interest in the nation's well-being consults widely and protects civil liberties generally, but in the end, decides on a course of action in the decisionmaker's sole discretion, unchecked by any other institutions.”); Balkin, *supra* note 13, at 2024;

to Zuckerberg's unchecked power over what Zuckerberg calls the Facebook "community."³³ The introduction of the FOB therefore raises the same question that comparative constitutional literature has grappled with: why do dictators allow courts any degree of judicial independence which might interfere with their power?³⁴ To be clear, the use of the term "authoritarian" is descriptive and without normative judgment. Facebook is a company, not a nation state.³⁵ Authoritarianism in the sense of an executive having decision-making capacity unbounded by formal internal checks is not unusual in this context. But the analogy is helpful in understanding why Facebook might introduce an "independent" check and balance into its governance.³⁶

After briefly sketching the impossible challenge of content moderation that the FOB is intended to help solve, this section draws on the rich literature on courts in authoritarian regimes to examine the reasons the FOB, and the renunciation of power that it represents, might appeal to Facebook as a solution.

A. Facebook's Approach to the Impossible Task of Content Moderation

Consistent and coherent content moderation on the scale of a platform like Facebook is essentially an impossible challenge.

Henry Farrell et al., *Mark Zuckerberg Runs a Nation-State, and He's the King*, VOX (Apr. 9, 2018), <https://www.vox.com/the-big-idea/2018/4/9/17214752/zuckerberg-facebook-power-regulation-data-privacy-control-political-theory-data-breach-king> [https://perma.cc/W2Z3-EQED]; MACKINNON, *supra* note 4; Chander, *supra* note 4.

³³ Justin Hendrix, *Is Facebook a Community? Digital Experts Weigh In*, JUST SECURITY (Apr. 13, 2018), <https://www.justsecurity.org/54850/facebook-community-digital-experts-weigh/> [https://perma.cc/2UNT-JQY5].

³⁴ See generally MEHDI SHADMEHR ET AL., JUDICIAL INDEPENDENCE AND HUMAN RIGHTS IN AUTOCRACIES (2019); RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES (Tamir Moustafa & Tom Ginsburg eds., 2008).

³⁵ ANDREW KEANE WOODS, TECH FIRMS ARE NOT SOVEREIGNS (2018).

³⁶ See, e.g., Tamir Moustafa, *Law and Courts in Authoritarian Regimes*, 10 ANN. REV. L. & SOC. SCI. 281 (2014) (discussing the rich literature of the dynamics of courts in authoritarian regimes that helps explain how such institutions can further an autocrat's aims).

Facebook has over 2 billion monthly active users,³⁷ and over 2.5 billion pieces of content are shared on its platform every day.³⁸ This content is moderated to align with Facebook's public rules, called "Community Standards,"³⁹ and its internal guidelines.⁴⁰ In the first quarter of 2019 alone, Facebook "took action" on nearly 1.873 billion pieces of content for being in breach of these rules.⁴¹ Excluding the 1.8 billion pieces of content Facebook marked as spam, this means Facebook classified nearly 73 million pieces of content as falling within the other categories in its Community Standards: adult nudity and sexual activity, bullying and harassment, child nudity and sexual exploitation of children, hate speech, regulated goods, global terrorist propaganda, and violence and graphic content.⁴² Because not every piece of content flagged for review is actually found to violate Facebook's rules, these numbers represent only a subset of flagged items that Facebook evaluates.⁴³ Facebook receives more than a million reports of violations of its content standards *per day*.⁴⁴ As well as reviewing these reports, Facebook also re-reviews decisions that are appealed by users. In the first quarter of 2019, Facebook received nearly 25 million requests for appeal—around 275,000 requests per day.⁴⁵ Clearly, this does not give the 30,000 people Facebook currently employs to enforce its Community Standards much time to consider each decision.⁴⁶

³⁷ FACEBOOK NEWSROOM, *supra* note 10.

³⁸ GILLESPIE, *supra* note 29, at 114.

³⁹ *Community Standards*, *supra* note 25.

⁴⁰ Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1638–9 (2018).

⁴¹ *Community Standards*, *supra* note 25.

⁴² *Id.*

⁴³ See GILLESPIE, *supra* note 29, at 74.

⁴⁴ Bickert, *supra* note 31, at 256.

⁴⁵ *Community Standards*, *supra* note 25 (calculations on file with the author).

⁴⁶ Zuckerberg, *supra* note 1. For insight into the tough working conditions of front line content moderators, see SARAH T. ROBERTS, *BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA* (2019).

Simply employing more people would not solve Facebook's moderation woes (although it could certainly help).⁴⁷ While a larger workforce might be able to more carefully consider each decision (for example, by looking at more contextual information that might shed light on the intended meaning of a particular post), this might come at the cost of greater consistency between decisions.⁴⁸ Consistency is an important indicator of fairness. Because of this, and because of the sheer number of content moderation decisions Facebook has to make, the company has developed an "industrial" approach to content moderation.⁴⁹ The goal is to create a "decision factory," where application of Community Standards is reduced to bright-line rules, whose application is routinized and efficient.⁵⁰ This attempted clean-cut approach has attracted controversy, such as over the strict use of "protected categories" in Facebook's hate speech policy leading to posts attacking "white men" being found to contravene the Standards but not those targeting "black children."⁵¹

Another tool Facebook uses to make content moderation at scale manageable is artificial intelligence (AI).⁵² In his Blueprint, Zuckerberg calls AI "the single most important improvement in enforcing our policies," because it can quickly and proactively identify harmful content.⁵³ This further reflects Facebook's

⁴⁷ Ellen Silver, *Hard Questions: Who Reviews Objectionable Content on Facebook — And Is the Company Doing Enough to Support Them?*, FACEBOOK (July 26, 2018) <https://newsroom.fb.com/news/2018/07/hard-questions-content-reviewers/> [<https://perma.cc/2WGP-VPHQ>] (explaining that Facebook itself has recognized this, doubling its content moderation staff in 2018).

⁴⁸ ROBYN CAPLAN, DATA & SOCIETY, CONTENT OR CONTEXT MODERATION? 6, 13 (2018).

⁴⁹ GILLESPIE, *supra* note 29, at 77.

⁵⁰ Caplan, *supra* note 48, at 24.

⁵¹ Julia Angwin & Hannes Grassegger, *Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children*, PROPUBLICA (June 28, 2017), <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms> [<https://perma.cc/P4SU-Z7YK>].

⁵² Drew Harwell, *AI Will Solve Facebook's Most Vexing Problems, Mark Zuckerberg Says. Just Don't Ask When or How.*, WASH. POST (Apr. 11, 2018), <https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/> [<https://perma.cc/C3X6-2QFP>].

⁵³ Zuckerberg, *supra* note 1.

mechanistic approach to content moderation. For every category except bullying and harassment, and hate speech, Facebook found over 95% of the content it took down as violating its Community Standards before it was reported by a user, in large part because of its AI. But the exception of bullying and harassment, and hate speech is telling: these two categories of content are harder for Facebook to proactively identify because they are so highly context-dependent. As Facebook notes, bullying and harassment reports “tend to be personal and context-specific, so in many instances we need a person to report this behavior to us before we can identify or remove it. This results in a lower proactive detection rate than other types of violations.”⁵⁴ Hate speech is notoriously difficult to detect through automated processes, because it depends so much on linguistic nuance, intention, and local norms.⁵⁵ Context is all-important at these “complex frontiers of political speech, dangerous speech, and hate speech.”⁵⁶ So while AI is a necessary part of content moderation at scale, it is not sufficient.

Even if it was technologically possible to train AI to appreciate the infinite spectrum of human nuance, there are at least two more reasons why AI cannot be a complete answer to the content moderation problem.

First, AI does not give a person who has a decision made against them the substance or feeling of due process or of being heard.⁵⁷ AI

⁵⁴ Guy Rosen, *How Are We Doing at Enforcing Our Community Standards?*, FACEBOOK NEWSROOM (Nov. 15, 2018), <https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/> [<https://perma.cc/72PF-T6FG>].

⁵⁵ Evelyn Douek, *Zuckerberg's New Hate Speech Plan: Out With the Court and In With the Code*, LAWFARE (Apr. 14, 2018), <https://www.lawfareblog.com/zuckerbergs-new-hate-speech-plan-out-court-and-code> [<https://perma.cc/4HDT-PASA>] (noting Zuckerberg acknowledges the difficulty of using AI to detect hate speech in Congressional testimony, and discussing why it is difficult); Zuckerberg also acknowledged as much in his Blueprint. Zuckerberg, *supra* note 1 (“As you get into hate speech and bullying, linguistic nuances get even harder . . . the state of the art in AI is still not sufficient to handle these challenges on its own. So we use computers for what they’re good at -- making basic judgements on large amounts of content quickly -- and we rely on people for making more complex and nuanced judgements that require deeper expertise.”).

⁵⁶ ASH ET AL., *supra* note 16, at 11.

⁵⁷ See generally Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2007).

also does not give public reasoning for its decisions, which, as argued below,⁵⁸ is an essential component of a legitimate content moderation system.

Second, and even more fundamentally, AI cannot be a full answer to the difficulties of content moderation because AI cannot determine the anterior question of the values that should be encoded into the detection algorithms. Before asking AI to identify and remove hate speech from the platform, for example, it has to be told what to look for. But what constitutes impermissible hate speech is an essentially contested concept that varies around the world.⁵⁹ As one Facebook representative writes, “[t]here is no universally accepted answer for when something crosses the line. Although a number of countries have laws against hate speech, their definitions of it vary significantly.”⁶⁰ A well-known example is Holocaust denial. As a fairly concrete category of speech, less context dependent than many others, this might be a prime example of speech that AI might be able to more easily identify. But while Holocaust denial is illegal in several countries,⁶¹ it is famously not in the United States where the decision that Nazis should be allowed to march in front of Holocaust survivors is seen as one of the “truly great victories” in American legal history.⁶² AI itself cannot choose between these two opposing conceptions of free speech: it has to be told whether to find and remove Holocaust denial by a human. Another example is Facebook’s exception to its Community

⁵⁸ See Part V(B).

⁵⁹ See, e.g., Robert Post, *Hate Speech*, in EXTREME SPEECH AND DEMOCRACY 123 (Ivan Hare & James Weinstein eds., 2010); RONALD J. KROTOSZYNSKI, THE FIRST AMENDMENT IN CROSS-CULTURAL PERSPECTIVE: A COMPARATIVE LEGAL ANALYSIS OF THE FREEDOM OF SPEECH (2006); Sarah H. Cleveland, *Hate Speech at Home and Abroad*, in THE FREE SPEECH CENTURY 210 (Lee C. Bollinger & Geoffrey R. Stone eds., 2018).

⁶⁰ Richard Allan, *Hard Questions: Who Should Decide What Is Hate Speech in an Online Global Community?*, FACEBOOK NEWSROOM (June 27, 2017), <https://newsroom.fb.com/news/2017/06/hard-questions-hate-speech/> [https://perma.cc/LB5P-8QR9].

⁶¹ MICHAEL BAZYLER, HOLOCAUST, GENOCIDE, AND THE LAW: A QUEST FOR JUSTICE IN A POST-HOLOCAUST WORLD 183–209 (2017).

⁶² Geoffrey R. Stone, *Remembering the Nazis in Skokie*, HUFFINGTON POST (May 20, 2009), https://www.huffingtonpost.com/geoffrey-r-stone/remembering-the-nazis-in_b_188739.html. [perma.cc/H22C-3NLK].

Standards for content it deems to be “newsworthy.”⁶³ These cases require a balancing of the harm caused by allowing speech that breaches Facebook’s rules to remain on the platform against the public interest in being informed about the particular matter. Such balancing cannot be done in the abstract or ex ante and encoded into an algorithm, but requires case-by-case consideration.

Enter Facebook’s “Oversight Board,” which could temper the industrial application of content moderation rules at scale. The FOB can give a user an opportunity to be heard, and will then consider all the relevant context and competing values at stake in the case at hand. But more importantly, the FOB will offer an explanation for why content is or is not allowed on Facebook’s platform.

B. The Benefits of Voluntary Restraints

The question remains, however, why outsource this role to an *independent* body? Greater transparency and reason-giving could be provided by Facebook employees and policy-makers within the current content moderation ecosystem. And despite growing calls,⁶⁴ Zuckerberg is showing no signs of wanting to relinquish any of his enormously powerful roles at his company.⁶⁵ His enjoyment of this power is a matter of legend.⁶⁶

⁶³ KATE KLONICK, KNIGHT FIRST AMEND. INST. EMERGING THREATS SERIES, FACEBOOK V. SULLIVAN (2018).

⁶⁴ Margaret Sullivan, *Embattled and In Over His Head, Mark Zuckerberg Should—At Least—Step Down As Facebook Chairman*, WASH. POST (Nov. 19, 2018), https://www.washingtonpost.com/lifestyle/style/embattled-and-in-over-his-head-mark-zuckerberg-should--at-least--step-down-as-facebook-chairman/2018/11/19/de00c2d8-ebfa-11e8-96d4-0d23f2aaad09_story.html. [perma.cc/VA7F-CSQ8].

⁶⁵ Kara Swisher, *Full Transcript: Facebook CEO Mark Zuckerberg on Recode Decode*, RECODE (July 18, 2018), <https://www.recode.net/2018/7/18/17575158/mark-zuckerberg-facebook-interview-full-transcript-kara-swisher> [perma.cc/KR44-C4S3] (confirming he has no intention of firing himself); Richard Lawler, *Zuckerberg Says Stepping Down at Facebook is “Not the Plan,”* ENGADGET, (Nov. 20, 2018) <https://www.engadget.com/2018/11/20/zuckerberg-says-stepping-down-at-facebook-is-not-the-plan/> [perma.cc/2YB3-J8VX] (confirming he has no intention of stepping down as Chairman of the board).

⁶⁶ Zuckerberg famously had business cards that declared “I’m CEO, Bitch.” MG Siegler, *Card Designer: The Inspiration For Zuckerberg’s “I’m CEO, Bitch”?* Steve Jobs., TECHCRUNCH (2010), <http://social.techcrunch.com/2011/>

Nevertheless, the establishment of the FOB does constitute a renunciation of a degree of Facebook's power. The FOB will have the power to reverse Facebook's decisions,⁶⁷ constraining the otherwise plenary discretion that Facebook currently has over what appears on its platform.⁶⁸ Its establishment will recreate a system of separation of powers, where an independent judicial-style body will oversee the other branches of Facebook's content moderation: the legislative branch that writes Facebook's content moderation rules, and executive actors who implement these rules (with the help of AI). This is an unprecedented governance structure for a private company. But it is, of course, the dominant form of governance in nation states.

The matter should not be overstated—Zuckerberg is not recreating liberal democratic governance. He is not subjecting himself or his role to democratic accountability. But the FOB initiative is in keeping with Zuckerberg's long-standing pronouncements that Facebook is “more like a government than a traditional company.”⁶⁹ By initially referring to the FOB as a “Supreme Court” and calling its rules of operation a “Charter,” for example, Zuckerberg is implicitly distinguishing the FOB from other, more well-known forms of online dispute resolution (ODR) such as eBay's Resolution Center.⁷⁰ The message is that this is not ordinary commercial customer relations management; this is something grander. And the distinction is accurate. While most ODR systems are directed at resolving disputes between two private

06/25/im-ceo-bitch/ [perma.cc/3FCF-CSX4].

⁶⁷ DRAFT CHARTER, *supra* note 2, at 1.

⁶⁸ Klonick, *supra* note 40, at 1617 (“These platforms are both the architecture for publishing new speech and the architects of the institutional design that governs it. Because of the wide immunity granted by § 230, these architects are free to choose which values they want to protect — or to protect no values at all.”).

⁶⁹ DAVID KIRKPATRICK, THE FACEBOOK EFFECT: THE INSIDE STORY OF THE COMPANY THAT IS CONNECTING THE WORLD 254 (2010) (quoting Zuckerberg as saying, “[i]n a lot of ways Facebook is more like a government than a traditional company. We have this large community of people, and more than other technology companies we're really setting policies.”).

⁷⁰ ETHAN KATSH & ORNA RABINOVICH-EINY, DIGITAL JUSTICE: TECHNOLOGY AND THE INTERNET OF DISPUTES 34 (2017).

parties (such as a buyer and seller),⁷¹ the FOB will be resolving disputes more analogous to public law. The disputes it will hear will revolve around the exercise of power by the “government” of Facebook.⁷² In making decisions about the rules in what has become a key channel of public discourse, the FOB’s decisions will also need to take into account a broader conception of the “public interest,” rather than the mere resolution of the immediate dispute before it. So, while many of these other ODR systems are increasingly looking to automate processes and displace human dispute handlers,⁷³ the FOB looks to reintroduce and elevate the human component, recalling more conventional offline dispute resolution bodies.

There are four apparent reasons why this governance structure may appeal to Facebook for solving the content moderation dilemmas outlined above: (1) bestowing content-moderation decisions with an aura of legitimacy, aiding user-relations; (2) staving off or guiding more extensive governmental regulation; (3) outsourcing controversial decisions away from the company; and (4) facilitating better enforcement of existing standards. These closely track the reasons authoritarian regimes often retain court systems at least somewhat independent from the regime.⁷⁴

⁷¹ Rory Van Loo, *The Corporation as Courthouse*, 33 YALE J. ON REG. 57, 554 (2016); see also Josh Dzieza, *Dirty dealing in the \$175 billion Amazon Marketplace*, VERGE (Dec. 19, 2018), <https://www.theverge.com/2018/12/19/18140799/amazon-marketplace-scams-seller-court-appeal-reinstatement> [perma.cc/6BFB-MLPT].

⁷² *Public Law*, BLACK’S LAW DICTIONARY (10th ed. 2014) (Public law is defined as “[t]he body of law dealing with the relations between private individuals and the government, and with the structure and operation of the government itself”).

⁷³ KATSH & RABINOVICH-EINY, *supra* note 70, at 38.

⁷⁴ Moustafa and Ginsburg suggest five functions for courts in an authoritarian state: “to (1) establish social control and sideline political opponents, (2) bolster a regime’s claim to ‘legal’ legitimacy, (3) strengthen administrative compliance within the state’s own bureaucratic machinery and solve coordination problems among competing factions within the regime, (4) facilitate trade and investment, and (5) implement controversial policies so as to allow political distance from core elements of the regime.” Tamir Moustafa & Tom Ginsburg, *Introduction, in* RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES 4 (Tamir Moustafa & Tom Ginsburg eds., 2008).

1. *Bestowing Legitimacy and Reassuring Users*

Legitimacy is central to the FOB experiment. The topic of legitimacy is discussed at length below,⁷⁵ but two points are worth noting here. First, “legitimacy” can mean many things and is hard to define.⁷⁶ Throughout this paper, references to “legitimacy” are references to what Fallon calls “sociological legitimacy”—that is, the prevailing public attitude towards Facebook’s content moderation.⁷⁷ Importantly, legitimacy does *not* mean correctness; instead, “in circumstances of relatively widespread reasonable disagreement, . . . legitimacy connote[s] respect-worthiness.”⁷⁸ Therefore, the mission to legitimate Facebook’s content moderation ecosystem aims to create a situation where its decisions are generally considered somewhat worthy of respect, even if there are those that still disagree with the substance of those decisions. The second point worth noting is that this definition of legitimacy highlights the very low baseline from which Facebook is operating. In the past few years, Facebook’s decisions have increasingly come to be viewed as inconsistent and arbitrary, and therefore illegitimate.⁷⁹ A recent Pew survey found that only 31% of U.S. adults have a great deal or fair amount of confidence in social media companies to determine what offensive content should be removed from their platforms.⁸⁰ For the FOB to be successful in conferring legitimacy to Facebook’s content moderation, it does not need to create a situation where content moderation comes to be viewed as

⁷⁵ See *infra* Part IV.

⁷⁶ RICHARD H. FALLON, LAW AND LEGITIMACY IN THE SUPREME COURT 6 (2018).

⁷⁷ *Id.* at 21 (that is, this paper is not directly concerned with the legal or moral legitimacy of Facebook’s decisions, although these will of course affect sociological legitimacy).

⁷⁸ *Id.* at 8.

⁷⁹ ASH ET AL., *supra* note 16, at 18; Kaye, *supra* note 14; Nicolas Suzor et al., *Evaluating the Legitimacy of Platform Governance: A Review of Research and a Shared Research Agenda*, 80 INT’L COMM. GAZ. 385 (2018).

⁸⁰ John LaLoggia, *U.S. public has little confidence in social media companies to determine offensive content*, PEW RESEARCH CENTER, <https://www.pewresearch.org/fact-tank/2019/07/11/u-s-public-has-little-confidence-in-social-media-companies-to-determine-offensive-content/> [https://perma.cc/TB27-DH3T].

perfectly legitimate. In some ways, the FOB's job is much simpler than that: it only needs to create improvements over the current state of widespread perceptions of illegitimacy.

Independent, court-like institutions are one tool for mitigating the extent to which an otherwise unconstrained ruler's decisions are viewed as wholly illegitimate. As Dixon and Landau write, “[w]hen exercising powers of judicial review, most courts are . . . afforded a degree of presumptive legitimacy.”⁸¹ Authoritarian regimes use courts in this way: because they lack democratic procedural legitimacy in the exercise of power, such regimes often use courts to give a patina of more substantive legitimacy to their rule.⁸² Why should such regimes care about legitimacy in the first place? For authoritarian regimes, legitimacy is important because it can reduce the costs and necessity of resorting to force to maintain power.⁸³ For Facebook, the calculation is somewhat different but not entirely so. Because “code is law,”⁸⁴ Facebook's enforcement of its rules does not rely on force in the same way.⁸⁵ However, such enforcement will have other costs if stakeholders oppose Facebook's rulings. Facebook needs the approval of governments, users, advertisers, and the media in establishing its speech norms because these stakeholders can exert commercial pressure on the company—these are the “business reasons” that Facebook must pay attention to external reactions to how it polices its platform.⁸⁶

Indeed, because “exit” (i.e. leaving the platform) is easier than physical exit from a state, the costs of illegitimate decisions may be even greater. While network effects make it more unlikely that Facebook will become the next Myspace, a social media graveyard of abandoned profiles,⁸⁷ the last few years of scandals no doubt make

⁸¹ David Landau & Rosalind Dixon, *Abusive Judicial Review: Courts Against Democracy*, 53 U.C. DAVIS L. REV. ___, 23 (forthcoming 2020).

⁸² Moustafa, *supra* note 36, at 286.

⁸³ Moustafa & Ginsburg, *supra* note 74, at 5.

⁸⁴ LAWRENCE LESSIG, CODE V2.0 (Version 2.0 ed. 2006).

⁸⁵ JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 101–26 (2008).

⁸⁶ Bickert, *supra* note 31, at 265.

⁸⁷ See John Herrman, *What Happens When Facebook Goes the Way of Myspace?*, N.Y. TIMES (Dec. 12, 2018), <https://www.nytimes.com/2018/12/12/>

Facebook afraid to be complacent. As a Facebook-commissioned report by a group of independent academics explained:

Facebook has considerable capacity to manage its content regulation process in a top down manner which pays minimal attention to users' views. However, the very existence of the [Community Standards Enforcement Report] highlights the recognition that public views about Facebook and its attitude about the public matter. They matter for the individual user both because disgruntled users find ways to circumvent rules, for example opening multiple accounts. In addition, unhappy customers are less likely to use the site and more likely to seek alternatives to it.⁸⁸

Put simply, authoritarians and companies value legitimacy because it can reduce the costs of decisions with which people disagree.⁸⁹

Perceived legitimacy not only improves Facebook's relationship with its users, but also with its commercial partners such as advertisers. Such partners may be dissuaded from dealing with Facebook if it becomes too unpredictable or entails high reputational costs. Again, this may seem a distant threat given Facebook's practical duopoly over online advertising, but it no doubt still factors into Facebook's decisions. The FOB's role in mitigating this threat again resembles the way authoritarian regimes use courts to respond to foreign investor fears over an unpredictable local business environment.⁹⁰ As Moustafa observes, "[i]n the age of global competition for capital, it is difficult to find any government that is not engaged in some program of judicial reform designed to make legal institutions more effective, efficient, and predictable for the

\magazine/what-happens-when-facebook-goes-the-way-of-myspace.html
[<https://perma.cc/EXK3-TE77>].

⁸⁸ BEN BRADFORD ET AL., THE JUSTICE COLLABORATORY OF YALE LAW SCHOOL, REPORT OF THE FACEBOOK DATA TRANSPARENCY ADVISORY GROUP 39 (Apr. 2019).

⁸⁹ As Sarah T. Roberts notes, "content moderation is fundamentally a matter of brand protection for the firm." Sarah T. Roberts, *Digital Detritus: "Error" and the Logic of Opacity in Social Media Content Moderation*, FIRST MONDAY (Mar. 5, 2018), <https://firstmonday.org/ojs/index.php/fm/article/view/8283/6649> [<https://perma.cc/9UQG-PP3U>].

⁹⁰ Alec Stone Sweet, *Constitutional Courts*, in THE OXFORD HANDBOOK OF COMPARATIVE CONSTITUTIONAL LAW 816, 820 (2012).

purpose of attracting global capital.”⁹¹ For example, Egypt bolstered its court processes in the late-1950s and 1960s to stem the severe capital flight it was experiencing as investors became wary of the consolidation of unchecked power in the Nasser regime.⁹² For both users and commercial partners, then, an independent body overseeing content moderation can reassure these stakeholders, raising the value of the Facebook “product” of content moderation by making it appear more stable and less subject to Facebook’s whims.⁹³

Courts are also a relatively low-cost way of achieving these legitimacy gains. For many authoritarian regimes, “[g]ranting access to the courts was a concrete way to relieve political pressure without opening the political system.”⁹⁴ That is, a court-like check is a far more limited surrender of power, defined in scope, than wholesale reform of the policy-making institutions. Facebook may think that by surrendering some power over content moderation decisions to the FOB, it is bolstering its claim to run the community for the benefit of its users and thereby weakening calls for more extensive reforms to decision-making at the company.

2. *Staving Off or Guiding Regulation*

Facebook may view the establishment of the FOB as a way of forestalling more extensive governmental regulation or at least as an opportunity to shape the form that such regulation may take. As tech platforms’ power and the related number of controversies grow, there is a growing sense that new laws are “inevitable.”⁹⁵ Even

⁹¹ Moustafa, *supra* note 36, at 285.

⁹² Tamir Moustafa, *Law and Resistance in Authoritarian States: The Judicialization of Politics in Egypt*, in *RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES* 132, 133–36 (Tamir Moustafa & Tom Ginsburg eds., 2008); Mark Tushnet, *Authoritarian Constitutionalism*, 100 *CORNELL L. REV.* 391, 423 (2015).

⁹³ Georg Vanberg, *Establishing and Maintaining Judicial Independence*, in *THE OXFORD HANDBOOK OF LAW AND POLITICS* 99, 107 (Gregory A. Caldeira et al. eds., 2008).

⁹⁴ Moustafa, *supra* note 36, at 287.

⁹⁵ Matt O’Brien & Mary Clare Jalonick, *Is Facebook Regulation “Inevitable”? Not So Fast*, *U.S. NEWS & WORLD REPORT* (Apr. 11, 2018), <https://www.usnews.com/news/business/articles/2018-04-11/is-facebook-regulation-inevitable-not-so-fast> [<https://perma.cc/4NCM-GLHM>].

Zuckerberg himself is now calling for global regulation of “harmful content.”⁹⁶ In the absence of global regulation, an ad hoc patchwork of local regulations could be costlier for companies by requiring them to comply with multiple different sets of rules. Research shows that even relatively modest voluntary efforts by private firms to restrain their own behavior can stave off much more stringent public regulations.⁹⁷ A good example of this is ad transparency on social media platforms. Facebook, Google, and Twitter have all recently unveiled voluntary ad transparency measures.⁹⁸ Meanwhile, the proposed Honest Ads Act, which would compel these sorts of disclosures, has made little progress in Congress,⁹⁹ despite extensive reporting about weaknesses in Facebook’s ad transparency tools.¹⁰⁰ Another example is tech platforms entering into voluntary Code of

⁹⁶ Mark Zuckerberg, *Four Ideas to Regulate the Internet*, FACEBOOK NEWSROOM (Mar. 30, 2019), <https://newsroom.fb.com/news/2019/03/four-ideas-regulate-internet/> [<https://perma.cc/LSF5-PPS5>].

⁹⁷ Neil Malhotra et al., *Does Private Regulation Preempt Public Regulation?*, AM. POL. SCI. REV. 1 (2018).

⁹⁸ Issie Lapowsky & Louise Matsakis, *You Can Now See All the Ads Facebook Is Running Globally*, WIRED (June 28, 2018), <https://www.wired.com/story/facebook-aims-more-transparency-view-ads-feature/> [<https://perma.cc/6HGF-5VFC>].

⁹⁹ Honest Ads Act, S. 1989, 115th Cong. (2017-2018).

¹⁰⁰ Matthew Rosenberg, *Ad Tool Facebook Built to Fight Disinformation Doesn’t Work as Advertised*, N.Y. TIMES (July 25, 2019), <https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html> [<https://perma.cc/C5NB-8N2R>]; Ariana Tobin & Jeremy B. Merrill, *Facebook’s Screening for Political Ads Nabs News Sites Instead of Politicians*, PROPUBLICA (June 15, 2018), <https://www.propublica.org/article/facebook-new-screening-system-flags-the-wrong-ads-as-political> [<https://perma.cc/7DZ7-NJES>]; April Glaser, *It’s Impossible to Tell if Russians Are Pushing Political Propaganda on Facebook*, SLATE MAGAZINE (July 5, 2018), <https://slate.com/technology/2018/07/facebooks-ad-transparency-system-isnt-nearly-good-enough.html> [<https://perma.cc/D9F3-23MD>]; J. Nathan Matias, Austin Hopkins & Melissa Hounsel, *We Tested Facebook’s Ad Screeners and Some Were Too Strict*, ATLANTIC (Nov. 2, 2018), <https://www.theatlantic.com/technology/archive/2018/11/do-big-social-media-platforms-have-effective-ad-policies/574609/> [<https://perma.cc/2G7N-PPS3>]; Ali Breland, *Dems push Facebook to fix “unacceptable” political ad transparency tools*, HILL (Nov. 2, 2018), <https://thehill.com/policy/technology/414524-dem-senators-push-facebook-to-clean-up-political-ad-transparency-loopholes> [<https://perma.cc/U6DE-YZ7M>].

Conduct agreements with the European Union, which held off coercive measures.¹⁰¹

Voluntary reforms can benefit governments too. Effectively regulating tech companies will be extremely difficult, and if done poorly, governments may be left taking the blame for poor outcomes. Self-regulatory reform at Facebook may make a lack of governmental regulation more politically tenable for politicians by showing some form of progress. Authoritarian states use formal compliance with constitutional requirements in this way, to help reduce the political costs for other states to maintain alliances.¹⁰² A recent report commissioned by the French government shows this dynamic playing out, with Facebook's voluntary reforms influencing the model of regulation ultimately endorsed by the report's authors. They recommended a model that focused on "expanding and legitimizing" platform self-regulation based on "the progress made in the last 12 months by . . . Facebook,"¹⁰³ showing that self-imposed platform reform and regulatory reform occur in dialogue with one another.¹⁰⁴

In some countries, the reform may be more anticipatory than others. In the U.S., the FOB might be seen as an attempt to preempt calls to narrow intermediary immunity provisions to deny platforms protection if they fail to take reasonable steps to prevent their services being used maliciously.¹⁰⁵ In other countries, the reform may merely be responsive to certain regulations that have already been put in place. A notable example is recent regulation in Germany that Facebook's head of policy for Europe has described as already "pushing us to the role of the court, the role of the legal

¹⁰¹ Citron, *supra* note 13, at 1041–44.

¹⁰² Moustafa & Ginsburg, *supra* note 74, at 6 (giving the examples of postwar Korea and Taiwan, and the Philippines under Ferdinand Marcos).

¹⁰³ FRENCH SEC'Y OF STATE FOR DIGITAL AFFAIRS, REGULATION OF SOCIAL NETWORKS – FACEBOOK EXPERIMENT (May 2019).

¹⁰⁴ For more on this relationship between voluntary initiatives and government regulation, see Emily B. Laidlaw, REGULATING SPEECH IN CYBERSPACE: GATEKEEPERS, HUMAN RIGHTS AND CORPORATE SOCIAL RESPONSIBILITY 78-83 (2015).

¹⁰⁵ Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Sec. 230 Immunity*, 86 FORDHAM L. REV. 401 (2017).

system.”¹⁰⁶ In either case, Facebook’s establishment of the FOB makes it not merely a passive recipient of regulatory mandates but a proactive player in the design of the future of Internet governance.

3. *Outsourcing Controversy*

For many of the issues that arise in the course of content moderation, there will be no “right” answer and any decision is likely to upset a certain constituency. The FOB allows Facebook to pass the responsibility for these divisive decisions to an independent body: a renunciation of power in the hope of also outsourcing some of the blame for contentious choices.

A desire for Facebook to no longer bear the brunt of public opprobrium for content moderation decisions is evident in Zuckerberg’s call for third-party bodies to set global standards for “harmful content.”¹⁰⁷ A more independent body set up by regulators would further distance Facebook from these choices, but global cooperation is unlikely to be forthcoming in the near future, if ever. Setting up the FOB might be seen as the next-best alternative. Democratic¹⁰⁸ and authoritarian regimes¹⁰⁹ alike use courts as a shield for controversy that can attend divisive political decisions.¹¹⁰ Despite having “no influence over either the sword or the purse,”¹¹¹

¹⁰⁶ Max Fisher & Amanda Taub, *How Far Should Facebook Go in Regulating Users’ Speech?*, N.Y. TIMES (Jan. 10, 2019), https://static.nytimes.com/email-content/INT_9276.html?nlid=78801897 [<https://perma.cc/CM6N-ML79>].

¹⁰⁷ Zuckerberg, *supra* note 96.

¹⁰⁸ GEORGE I. LOVELL, LEGISLATIVE DEFERRALS: STATUTORY AMBIGUITY, JUDICIAL POWER, AND AMERICAN DEMOCRACY (2003); Keith E. Whittington, “Interpose Your Friendly Hand:” *Political Supports for the Exercise of Judicial Review by the United States Supreme Court*, 99 AM. POL. SCI. REV. 583 (2005).

¹⁰⁹ Moustafa, *supra* note 36, at 286.

¹¹⁰ Mark A. Graber, *The Nonmajoritarian Difficulty: Legislative Deference to the Judiciary*, 7 STUD. AM. POL. DEV. 35, 43 (1993) (“[T]he aim of legislative deference to the judiciary is for the courts to make controversial policies that political elites approve of but cannot publicly champion, and to do so in such a way that these elites are not held accountable by the general public, or at least not as accountable as they would be had they personally voted for that policy[.]”); Vanberg, *supra* note 93, at 108 (“Rather than make decisions that carry an electoral risk, officials may prefer to pass such issues off to someone else, including the judiciary. However, this strategy of ‘passing the buck’ only works if the other actor is independent.”).

¹¹¹ THE FEDERALIST NO. 78 (Alexander Hamilton).

courts are obeyed by other governmental actors at least in part because “[d]isaffected citizens can then blame the [courts], not elected officials, for their disappointments.”¹¹²

The motivation to avert blame is not limited to governments—other actors can also use the law to avoid social backlash for their decisions. For example, many voluntarily segregated restaurants and hotels lobbied for civil rights legislation that would have forced them to desegregate.¹¹³ Segregation was commercially costly for hotels, depriving them of black customers, but voluntarily desegregating would have imposed a social cost in their prejudiced communities. Civil rights legislation made desegregation a matter of legal compliance rather than a decision with social meaning, and outsourced the value judgment.¹¹⁴

Similarly, when the FOB decides that a certain popular commentator should be banned for sprouting hate speech or that certain misleading political content should be left up rather than completely removed, Facebook can cast implementing these decisions as mere compliance rather than its own politically-charged decisions. This fits with Facebook’s own professed image as apolitical, and agnostic as to the ideology of content on its platform.¹¹⁵ Facebook does not want to take sides in the culture wars, presumably because doing so would alienate segments of its

¹¹² FALLON, *supra* note 76, at 115.

¹¹³ Cass R. Sunstein, *Unleashed*, 85 SOC. RES. 73 (2017); Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. CHI. L. REV. 943, 966 (1995).

¹¹⁴ Sunstein, *supra* note 113, at 79 (“[T]hey wanted not to discriminate, because discrimination on their part was costly. But in light of prevailing norms, they would also have incurred a high cost for not discriminating, which would have provoked a hostile reaction in their community. As Lawrence Lessig writes, ‘for a white to serve or hire blacks was for the white to mark him or herself as having either a special greed for money or a special affection for blacks.’ In these circumstances, the force of the law was needed to alter the social meaning of non-discrimination. Once the Civil Rights Act of 1964 was enacted, non-discrimination was a matter of compliance. Profit-making companies were liberated.”) (citations omitted).

¹¹⁵ See Nick Statt, *Zuckerberg calls Facebook ‘a platform for all ideas’ after meeting with conservatives*, VERGE (May 18, 2016), <https://www.theverge.com/2016/5/18/11706266/mark-zuckerberg-facebook-conservative-news-censorship-meeting> [<https://perma.cc/9PFE-2GXS>].

customer base. After all, Republicans buy shoes from Facebook ads too.¹¹⁶

Of course, there are limits to the extent to which Facebook can avoid responsibility for what it allows on its platform. As one representative has written, “[t]o be clear, we are not asking a group of experts to make decisions for us. We are, however, asking for their insights to help inform our thinking and hold us accountable. We’ll still be making hard decisions every day, and we accept the full weight of that responsibility.”¹¹⁷ Facebook cannot be seen to be trying to wash its hands of the problems its platform creates, nor does it want to completely renounce control over important content decisions that define the user experience. Finding the right balance will no doubt involve a degree of ongoing calibration.

4. *Enforcing Existing Standards*

To a limited extent, the FOB might also help not only with changing or filling in gaps in Facebook’s policies, but also with enforcing the policies that Facebook already has. In any complex system, enforcement error is inevitable. As Zuckerberg says in his Blueprint, “[t]he vast majority of mistakes we make are due to errors enforcing the nuances of our policies rather than disagreements about what those policies should actually be.”¹¹⁸ An appeals body can be a useful mechanism for monitoring the performance of and correcting errors in the application of its existing standards. It has the advantage of distributing the burden of finding mistakes by making private parties bear the cost of rectifying errors by bringing cases. This is how many authoritarian regimes use a system of administrative law courts, to resolve principal-agent problems in the

¹¹⁶ See Laura Wagner, “Republicans Buy Sneakers, Too,” SLATE (July 28, 2016), <https://slate.com/culture/2016/07/did-michael-jordan-really-say-republicans-buy-sneakers-too.html> [<https://perma.cc/6B2M-64FG>].

¹¹⁷ Ime Archibong, *Strengthening Safety and Encouraging Free Expression: Facebook’s Goals for a New Content Oversight Board*, FACEBOOK NEWSROOM (Mar. 26, 2019), <https://newsroom.fb.com/news/2019/03/ime-op-ed/> [<https://perma.cc/6D52-PHBY>].

¹¹⁸ Zuckerberg, *supra* note 1.

administration of their policies that cannot be resolved through centralized monitoring alone.¹¹⁹

Facebook may, for example, be satisfied that it has drawn the appropriate line in its definition of hate speech.¹²⁰ But this highly contextual and sensitive judgment call will not always be implemented correctly by context-blind AI tools or the burnt-out, time-pressured, and geographically-distributed contractors that are charged with implementing these rules.¹²¹ Currently, it is not uncommon for external parties to find such errors and to extract apologies from Facebook once they draw public attention to the mistakes.¹²² Creating the FOB formalizes this process and redirects the public outrage into a predetermined process for the resolution of disputes. This has the added benefit of encouraging *engagement* with the platform and its rules, rather than mere criticism or flight.¹²³

¹¹⁹ See Tom Ginsburg, *Administrative Law and the Judicial Control of Agents in Authoritarian Regimes*, in RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES 58, 59 (Tamir Moustafa & Tom Ginsburg eds., 2008) (“To resolve the problem and prevent agency slack, all rulers need mechanisms to monitor agents’ performance and to discipline agents who do not obey instructions.”); Moustafa, *supra* note 92, at 140–41.

¹²⁰ Facebook defines hate speech as “a direct attack on people based on what we call protected characteristics—race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity and serious disease or disability. We also provide some protections for immigration status. We define ‘attack’ as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.” FACEBOOK, *supra* note 25.

¹²¹ See Roberts, *supra* note 20; see also THE CLEANERS documentary.

¹²² See, e.g., Ariana Tobin et al., *Facebook’s Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up*, PROPUBLICA (Dec. 28, 2017), <https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes> [<https://perma.cc/DY23-WJGA>]; Amy Pollard, *Facebook Found “Hate Speech” in the Declaration of Independence*, SLATE (July 5, 2018), <https://slate.com/technology/2018/07/facebook-found-hate-speech-in-the-declaration-of-independence.html> [<https://perma.cc/2DKV-GKG5>]; Sam Levin et al., *Facebook Backs Down From “Napalm Girl” Censorship and Reinstates Photo*, GUARDIAN (Sept. 9, 2016), <http://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo> [<https://perma.cc/CB2M-ZMMM>].

¹²³ See STONE SWEET, *supra* note 8, at 141 for a similar interpretation of the exercise of delegation of disputes to a court as a legitimation.

This process of engagement keeps users invested in the site and creates a kind of buy-in.

But it is important to be realistic about the extent of this function for the FOB: the FOB will only be able to hear and decide appeals in the tiniest fraction of content moderation decisions Facebook makes. Error correction cannot be its primary goal; it will be a collateral benefit only. And the extent to which regulators are persuaded to stay their hand and the public no longer blames Facebook alone for controversial decisions is dependent on whether the FOB itself is seen as legitimate and independent. Legitimacy is therefore the crux of the FOB experiment. Before discussing the fundamental challenge of creating legitimacy, the next section sketches the broad strokes of the FOB's institutional design.

III. THE FOB BLUEPRINT

Many details about the final institutional design of the FOB remain unknown at the time of writing.¹²⁴ Zuckerberg's Blueprint and the later released Draft Charter literally ask more questions than they answer. In his blog post, Zuckerberg asked: "how are members of the body selected? How do we ensure their independence from Facebook, but also their commitment to the principles they must uphold? How do people petition this body? How does the body pick which cases to hear from potentially millions of requests?"¹²⁵ As one report put it, "[a]ll the major questions remain unanswered."¹²⁶ The Draft Charter starts to fill in some details on the blueprint, but

¹²⁴ This section draws on Evelyn Douek, *Facebook's 'Draft Charter' for Content Moderation: Vague, But Promising*, LAWFARE (Jan. 31, 2019), <https://www.lawfareblog.com/facebooks-draft-charter-content-moderation-vague-promising> [<https://perma.cc/46NF-8PQU>]; see also Evelyn Douek & Kate Klonick, *Facebook Releases an Update on Its Oversight Board: Many Questions, Few Answers*, LAWFARE (June 27, 2019), <https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers> [<https://perma.cc/546D-T38A>]; Evelyn Douek, *Verified Accountability: Self-Regulation of Content Moderation as an Answer to the Special Problems of Speech Governance*, HOOVER AEGIS SERIES NO. 1903 (2019). This article reflects publicly available information about the FOB as of September 2019.

¹²⁵ Zuckerberg, *supra* note 1.

¹²⁶ ASH ET AL., *supra* note 16, at 20.

itself acknowledges that it is only a “starting point.”¹²⁷ Nevertheless, these documents do provide a rough picture of what the FOB will look like.

Throughout the development process, one noticeable shift has been the abandonment of the comparison of the FOB to the “Supreme Court.” This is welcome, and this paper argues that the ultimate design should go even further in this direction. Some scholars have suggested that the effectiveness of the FOB “will depend on the answer to one key question: How much will the ‘Supreme Court of Facebook’ be like the Supreme Court of the United States?”¹²⁸ Of course, the recourse of commentators and Zuckerberg to the analogy of the Supreme Court of the United States is not surprising. Facebook is an intrinsically American company, and its staff are acculturated in American culture and legal norms.¹²⁹ But focusing on a comparison to the U.S. legal system would be an unduly narrow view. Facebook is a global company. What’s more, since the U.S. Constitution was written, nearly 1,000 different constitutional systems have been created from which much has been learned.¹³⁰ Some of the details in the Draft Charter, described below, show Facebook is open to these lessons and to institutional innovation.

A. Membership

Ultimately, the individuals selected for the FOB will be enormously consequential for the quality of its decisions and the success of the project. But members will be helped or hindered by their institutional context. The Draft Charter announces that the FOB will be comprised of “up to 40 global experts,” selected on the basis of publicly available qualifications as well as geographical,

¹²⁷ DRAFT CHARTER, *supra* note 2.

¹²⁸ Kate Klonick & Thomas Kadri, *How to Make Facebook’s ‘Supreme Court’ Work*, N.Y. TIMES (Nov. 18, 2018), <https://www.nytimes.com/2018/11/17/opinion/facebook-supreme-court-speech.html> [<https://perma.cc/8YTD-C54Q>].

¹²⁹ Klonick, *supra* note 40, at 1621.

¹³⁰ ZACHARY ELKINS ET AL., *THE ENDURANCE OF NATIONAL CONSTITUTIONS* 6 (2009).

cultural, personal, and professional diversity.¹³¹ Deciding the size of the board involves a trade-off: a smaller group of members would create a greater concentration of expertise and authority, while a larger group would be more diverse and possibly able to hear more cases. Choosing to have members from varied professional backgrounds also shows this preference for diverse perspectives, instead of selecting members who would have a common set of professional norms and a shared discursive toolbox.

The choice not to prioritize accumulation of expertise and authority will be exacerbated by Board members' relatively short terms—fixed at three years and automatically renewable once, subject to removal for violation of terms of their appointment.¹³² Compensation will be set in advance and unchangeable.¹³³ The fixed tenure and compensation are important—these indicia are “the gold standard of independence” and essential to preserve members' independence from Facebook.¹³⁴ But the short term length does undermine this somewhat. Perhaps life tenure (as for U.S. federal judges and justices) is not necessary, feasible, or desirable for Facebook, but the international norm for judicial office on final courts of appeal is terms of around nine to fourteen years.¹³⁵ The choice of a shorter term not only gives members less time to develop their understanding of their role and the Facebook content

¹³¹ As to professional diversity, the Draft Charter says “The board will be made of experts with experience in content, privacy, free expression, human rights, journalism, civil rights, safety and other relevant disciplines.” DRAFT CHARTER, *supra* note 2, at 1.

¹³² *Id.* at 2 (the final Charter allows for members' terms to be renewed twice).

¹³³ *Id.* at 4.

¹³⁴ THE FEDERALIST NO. 78 (Alexander Hamilton) (“[N]othing will contribute so much as [permanent tenure of judicial offices] to that independent spirit in the judges which must be essential to the faithful performance of so arduous a duty.”).

¹³⁵ See Mark Tushnet, *Judicial Accountability in Comparative Perspective*, in ACCOUNTABILITY IN THE CONTEMPORARY CONSTITUTION 57, 57, 61–62 (Nicholas Bamforth & Peter Leyland eds., 2013); David S. Law, *How To Rig the Federal Courts*, 99 GEO. L.J. 779, 786 (2011) (“It is perhaps typical to think of appointment for life and protection against diminution in salary as the gold standards of judicial independence.”); Frank Cross, *Judicial Independence*, in THE OXFORD HANDBOOK OF LAW AND POLITICS 567 (Gregory A. Caldeira et al. eds., 2008); Vanberg, *supra* note 93, at 101; Stone Sweet, *supra* note 90, at 824 (suggesting term of 9–12 years is typical).

moderation ecosystem, but may also mean FOB members are unduly concerned with their career prospects post-service.¹³⁶ This perverse incentive to consider their own interests might be further aggravated by the fact that members will only be serving in a part-time capacity.¹³⁷ Indeed, this was promoted as a benefit by Noah Feldman in an early white paper on the FOB, arguing “[r]eputational effects will be important to them. They will be mid-career actors, so that they will have other interests and goals alongside their Facebook judicial position.”¹³⁸ While it is true, as Feldman argues, that this will reduce pressure to break professional norms and engage in attention-seeking behavior, having shorter, part-time terms increases the salience of board members’ personal interests rather than their institutional responsibility. These concerns might be somewhat mitigated by the fact that “individual members’ names will not be associated with particular decisions.”¹³⁹ Granting anonymity may protect against pressure to vote in a particular way on any given case, and increase the desire to bolster the legitimacy of the institution as a whole with respect-worthy decisions.

Overall, these choices reflect a theme that dominates the Draft Charter: the importance of diversity. Indeed, the words “diverse” or “diversity” are used six times in two pages.¹⁴⁰ The reason for this is clear. Facebook has long been criticized for the lack of diversity in its leadership, product design, and content moderation teams.¹⁴¹ This

¹³⁶ Tushnet, *supra* note 135, at 61–62

([I]f reappointment is precluded, judges may shape their decisions with an eye to their prospects for post-service employment, which may promote indirect accountability to someone, though not necessarily the public. And, for obvious reasons, these effects may differ depending on the age at which judges take their positions on apex courts, and whether they are subject to rules requiring or norms encouraging retirement at a specific age or restrictions on post-service employment.).

¹³⁷ See DRAFT CHARTER, *supra* note 2, at 2.

¹³⁸ Brent Harris, *Global Feedback & Input on the Facebook Oversight Board for Content Decisions*, FACEBOOK NEWSROOM (June 27, 2019), <https://newsroom.fb.com/news/2019/06/global-feedback-on-oversight-board/> [<https://perma.cc/YUW9-SWUN>].

¹³⁹ DRAFT CHARTER, *supra* note 2, at 4.

¹⁴⁰ *Id.* at 1–2.

¹⁴¹ See, e.g., Jaime Dunaway, *Lawmakers Press Zuckerberg on Facebook's Diversity Problem—and He Had an Answer Ready*, SLATE (Apr. 24, 2018),

has affected the company's ability to anticipate and understand how its platform impacts its diverse global user base. Many of the biggest content moderation controversies have involved the harsher operation of Facebook's rules on minorities, for example.¹⁴² Greater diversity enhances the legitimacy of courts' decisions in the eyes of the public, and especially for those minorities that previously lacked representation.¹⁴³ Facebook has written that "[t]he success and the ultimate effectiveness of the Oversight Board will depend on its ability to accommodate an inclusive and diverse range of perspectives, across language, culture and experience."¹⁴⁴ And in a global consultation process on the FOB, "many stressed the need for the Board's composition to reflect as many segments of society as possible."¹⁴⁵

But this may set an impossible target. The FOB cannot encompass diversity sufficient to represent the views of a user base that is perhaps the most diverse "community" ever assembled. The

<https://slate.com/technology/2018/04/lawmakers-question-mark-zuckerberg-on-facebooks-employment-diversity.html> [<https://perma.cc/9ZE3-AWEW>]; Tom Simonite, *AI Is the Future—But Where Are the Women?*, WIRED (Aug. 17, 2018), <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance> [<https://perma.cc/A3MK-N22J>]; Dave Lee, *Who's Policing Facebook?*, BBC (May 22, 2018), <https://www.bbc.com/news/technology-44206996> [<https://perma.cc/AG68-RS49>].

¹⁴² See, e.g., Mara Schechter, *Update: 570k+ People Call on Facebook to Stop Censoring Activists*, DAILY KOS (Mar. 3, 2017), <https://www.dailykos.com/story/2017/3/3/1639597/-Update-570k-people-call-on-Facebook-to-stop-censoring-activists> [<https://perma.cc/9X4H-QZ24>] (reporting on a letter from civil rights organizations, including the ACLU, criticizing Facebook: "Activists in the Movement for Black Lives have routinely reported the takedown of images discussing racism and during protests, with the justification that it violates Facebook's Community Standards. At the same time, harassment and threats directed at activists based on their race, religion, and sexual orientation is thriving on Facebook."); Angwin & Grassegger, *supra* note 51.

¹⁴³ See, e.g., Nancy Scherer, *Diversifying the Federal Bench: Is Universal Legitimacy for the U.S. Justice System Possible?*, 105 NW. U. L. REV. 587 (2011); Nancy Scherer & Brett Curry, *Does Descriptive Race Representation Enhance Institutional Legitimacy? The Case of the U.S. Courts*, 72 J. OF POLITICS 90 (2010); Stone Sweet, *supra* note 90 (noting also the importance of diversity of expertise).

¹⁴⁴ Archibong, *supra* note 117.

¹⁴⁵ Harris, *supra* note 138, at 20.

U.S. Supreme Court has been criticized for failing to reflect the diversity of its single-country jurisdiction.¹⁴⁶ Consider, then, that the FOB's jurisdiction is not a single country, but spans the globe. One report noted that "[c]reating a body that has credibility with the extraordinarily wide geographical, cultural, and political range of Facebook users would be a major challenge."¹⁴⁷ This is an understatement. It is not just that the FOB cannot hope to represent every single view, or cannot "include representatives from every country and culture."¹⁴⁸ It cannot hope to represent even a meaningful *fraction* of those views. This is especially so given that the forty-member FOB will not sit as a whole in each case, but will instead hear cases in panels.¹⁴⁹ A forty-member body is not even large enough to include a representative from each of Africa's fifty-four countries. Facebook acknowledges Africa is a "complex tapestry of cultures," but then proceeds to flatten this diversity by saying "Africa will be represented on the board alongside other regions."¹⁵⁰ There is no reason to believe, for example, that any representative from Africa will have the expertise and legitimacy necessary to issue authoritative pronouncements on issues as diverse as Nigerian fake news¹⁵¹ to Libyan "keyboard warriors" who carry out online partisan battles.¹⁵²

Rather than pretending that the FOB itself can be sufficiently diverse to be representative of Facebook users, the limits of this capacity should be factored into the FOB's design. There is no single

¹⁴⁶ Valerie Strauss, *The 'Cloistered' Harvard-Yale Law Monopoly on the Supreme Court*, WASH. POST (July 10, 2018), <https://www.washingtonpost.com/news/answer-sheet/wp/2018/07/10/the-cloistered-harvard-yale-law-monopoly-on-the-supreme-court/> [<https://perma.cc/QE4J-8UTZ>].

¹⁴⁷ ASH ET AL., *supra* note 16, at 20.

¹⁴⁸ Archibong, *supra* note 117.

¹⁴⁹ DRAFT CHARTER, *supra* note 2, at 3.

¹⁵⁰ Archibong, *supra* note 117.

¹⁵¹ Yemisi Adegoke et al., *Like. Share. Kill.*, BBC NEWS (Nov. 13, 2018), https://www.bbc.co.uk/news/resources/idx-sh/nigeria_fake_news?utm_campaign=The%20Interface&utm_medium=email&utm_source=Revue%20newsletter [<https://perma.cc/NZ2C-SGBC>].

¹⁵² Declan Walsh & Suliman Ali Zway, *A Facebook War: Libyans Battle on the Streets and on Screens*, N.Y. TIMES (Sept. 4, 2018), <https://www.nytimes.com/2018/09/04/world/middleeast/libya-facebook.html> [<https://perma.cc/8KVX-VXTJ>].

body that could represent a community the size and scale of Facebook. The FOB will instead need to be empowered to reach out to different experts and communities who have relevant perspectives to offer in any individual case. Facebook is not wrong that the legitimacy and therefore success of the FOB will depend in large part on its capacity to bring greater diversity to Facebook's content moderation system, but it will need to find ways to do this other than through FOB membership alone.

B. Power of Review

The Draft Charter states that the “primary function of the board is to review specific decisions we make when enforcing our Community Standards.”¹⁵³ As noted above, this may be the primary *activity* of the FOB, but it cannot be its primary *function*. Given the sheer volume of content moderation decisions that Facebook makes every day, the FOB's impact would be miniscule if its function were purely error correction in individual cases. Other aspects of the Draft Charter suggest Facebook appreciates this broader purpose.

Crucially, the FOB's decisions will be made public and include an explanation.¹⁵⁴ If the decision of the relevant panel is not unanimous, minority views can be included in the explanation for the decision.¹⁵⁵ This shows that the purpose of the FOB is not only to decide outcomes and correct errors, but also to give reasons—to expose the reasoning and tensions involved in content moderation decisions.

To guide this process, Facebook will publish a final Charter that includes a set of values which the FOB agrees to uphold.¹⁵⁶ This aims to make the basis for the FOB's decisions more transparent, thereby ensuring “the public legitimacy of the board will grow.”¹⁵⁷ FOB panels should “ensure consistency with other issued opinions”¹⁵⁸ so that its decisions create a coherent body of “platform

¹⁵³ DRAFT CHARTER, *supra* note 2, at 3.

¹⁵⁴ *Id.* at 5.

¹⁵⁵ *Id.*

¹⁵⁶ *Id.*

¹⁵⁷ *Id.*

¹⁵⁸ *Id.*

law.”¹⁵⁹ This emphasis on consistency and coherence with prior decisions invokes notions of *stare decisis*, the legal doctrine of precedent. Given this reliance on a fundamental concept of legal training, it is again notable that not all members of the FOB will be lawyers.

Given that reason-giving is central to the FOB’s role, its institutional design should be centered around facilitating the production of quality decisions. Again, this involves trade-offs, the most obvious being speed of decision-making in order to ensure that by the time the case is decided there is still the possibility of a substantive remedy. If a take-down decision comes too late, the damage of a viral post may already be done. The importance of speed versus careful consideration may differ in each case, but the Draft Charter indicates there will be a hard deadline of two weeks for all FOB decisions.¹⁶⁰ The difficulty of this rigid timeline is indicated by the wide range of activities that the FOB might undertake during this time: it will be empowered to “call upon experts to ensure it has all supplementary linguistic, cultural and sociopolitical expertise necessary to make a decision,”¹⁶¹ and other stakeholders will also be able to submit arguments and material to the panel.¹⁶² This additional, inquisitorial-style evidence gathering powers of the FOB are an essential departure from the U.S. Supreme Court model. The factual record on which Facebook’s content moderators base their decisions is incredibly slim—sometimes it does not even include comments or captions that accompany a piece of flagged content.¹⁶³ But correct content moderation decisions depend on a proper understanding of context. The same word could be a form of artistic expression in one context or a racial slur in another. Political, historical, or social context may make a phrase that seems innocuous on its face extremely provocative and even

¹⁵⁹ Kaye, *supra* note 14, at 3.

¹⁶⁰ DRAFT CHARTER, *supra* note 2, at 5. However, it is not clear from the Draft Charter when this clock will start running.

¹⁶¹ *Id.* at 4.

¹⁶² *Id.*

¹⁶³ LAURA MURPHY, FACEBOOK’S CIVIL RIGHTS AUDIT – PROGRESS REPORT 11 (June 30, 2019).

dangerous. Proper responses to hate and disinformation depends on an understanding of local, or even hyper-local, considerations.¹⁶⁴

Zuckerberg has also acknowledged that the “linguistic nuances” involved in identifying hate speech and bullying pose a special challenge for both reviewers and AI.¹⁶⁵ A civil rights audit found that a key reason for errors in Facebook’s moderation of hate speech was insufficient attention to context.¹⁶⁶ The FOB’s role in considering extra context and correcting errors will therefore be especially important in these cases.

In difficult cases, then, the extent of the material necessary for the FOB to make a well-informed decision could be vast, or the logistics of finding and hearing from necessary experts might be challenging. But because of the limitations on the FOB’s ability to adequately represent the full diversity of Facebook users, as noted above, this expertise-gathering should be given ample space to occur. In this context, a two-week deadline seems somewhat arbitrary: a functional board would itself determine the correct balance in each case. It is unclear why a strict standard of two weeks strikes a good balance—a fortnight is an eternity in terms of the Internet zeitgeist (justice delayed is virality denied) but perhaps not long enough for a multi-member board to gather and consider all the materials it needs.

Another notable departure from the U.S. model is the FOB being given a kind of “abstract” jurisdiction.¹⁶⁷ Facebook has indicated that beyond deciding individual cases, the company “may request policy guidance from the board.”¹⁶⁸ The U.S. Supreme Court’s jurisdiction hinges on the finding of a specific “case or

¹⁶⁴ CHINMAYI ARUN, REBALANCING REGULATION OF SPEECH: HYPER-LOCAL CONTENT ON GLOBAL WEB-BASED PLATFORMS (2018); David Kaye, *Four Questions About Regulating Online Hate Speech*, MEDIUM (Aug. 12, 2019), <https://onezero.medium.com/four-questions-about-online-hate-speech-ac3e0a134472> [<https://perma.cc/3D7Z-G6N6>].

¹⁶⁵ Zuckerberg, *supra* note 1.

¹⁶⁶ Murphy, *supra* note 163, at 11.

¹⁶⁷ Stone Sweet, *supra* note 90, at 818.

¹⁶⁸ DRAFT CHARTER, *supra* note 2, at 3.

controversy,”¹⁶⁹ but an abstract jurisdiction for the FOB will empower it to highlight problems in Facebook’s rules in advance of individuals bringing appeals about them. This capacity could be further enhanced by the creation of a concrete review jurisdiction,¹⁷⁰ where moderators might escalate matters to review in hard cases where they are uncertain about how the Community Standards should be applied, in advance of a user-initiated appeal. This would have the advantage of potentially heading off controversy created by moderators being forced to make a decision in hard cases without the benefit of full consideration. The power to assess the appropriateness of rules and not merely the application of them in a particular case is an important one, discussed further below, and should be available beyond Facebook’s discretion to refer such cases.

Many questions remain about how, in practice, the FOB will exercise these powers of review. Not least among them are “standing” rules about who is qualified to bring an appeal (just the user who posted the content, for example, or third parties?), the form of argumentation (will there be a Facebook bar who argues before the FOB in person?), how cases will be selected from the large volume of contested decisions, how to support “litigants” to ensure that access to the FOB is not limited to the well-resourced, who will bear the burden of proof to establish error, what kind of evidence the FOB will consider, and innumerable other questions. These decisions will impact the FOB’s legitimacy, and many will likely be worked out over time. This article is concerned with institutional design of the FOB at a higher level of generality, and so these questions are set aside for now. But once the FOB is operational, these questions will become more urgent and salient in how it is perceived.

C. Subject-Matter Jurisdiction

For the FOB to meaningfully contribute to Facebook’s content moderation ecosystem, its subject-matter jurisdiction—that is, the

¹⁶⁹ U.S. CONST., art. III, § 2, cl. 1. This is a common feature in the ideal type of a constitutional court. Stone Sweet, *supra* note 90, at 818.

¹⁷⁰ STONE SWEET, *supra* note 8, at 45.

topics on which it is empowered to hear cases—should reflect the wide ambit of ways Facebook decides what appears on its platform.

Two things are worth noting about the subject-matter jurisdiction of the FOB. First, the FOB “will not decide cases where reversing Facebook’s decision would violate the law.”¹⁷¹ This reflects Facebook’s general position that it will respect local laws in the countries it operates.¹⁷² Facebook follows local laws because they “are often the result of public input, even just the indirect influence of a democratic election. The laws are therefore likely to reflect, at least in democracies, the social values of the local population.”¹⁷³ This restriction on the FOB’s jurisdiction is not surprising: indeed, where local law conflicts with Facebook’s rules, the FOB would have limited power to mandate that Facebook disobey legal requirements. But a consequence of this is that governments have de facto control over the FOB’s remit. A government can prevent the FOB from giving an opinion on a matter if it makes a particular type of content illegal. As Feldman acknowledges, “[i]t would be plausible to eliminate this provision from the proposal altogether. The idea would be to discourage states from enacting laws that limit expression and hence *gaining control over content on Facebook*.”¹⁷⁴ This provision also undermines Facebook’s professed desire to have a universal set of Community Standards. In practice, and particularly in countries where governments seek to exercise tight control over public discourse, carving out disputes under local law may be a severe limitation. National and international laws about free speech and online content are proliferating and diverse.¹⁷⁵ Even within a single jurisdiction,

¹⁷¹ DRAFT CHARTER, *supra* note 2, at 3.

¹⁷² Bickert, *supra* note 31, at 258.

¹⁷³ *Id.*

¹⁷⁴ Harris, *supra* note 138, at 142; Noah Feldman, *Facebook Supreme Court: A Governance Solution*, in GLOBAL FEEDBACK AND INPUT ON THE FACEBOOK OVERSIGHT BOARD FOR CONTENT DECISIONS: APPENDIX 104, 106 (2019) (emphasis added).

¹⁷⁵ See, e.g., ERIC GOLDMAN, AN OVERVIEW OF THE UNITED STATES’ SECTION 230 INTERNET IMMUNITY, (Santa Clara Univ. Legal Studies Research Paper, 2018); see also Daniel Funke, *A guide to anti-misinformation actions around the world*, POYNTER (Jan. 8, 2019), <https://www.poynter.org/fact-checking/2019/a-guide-to-anti-misinformation-actions-around-the-world/>

different courts can come to different conclusions on the meaning of relevant laws and whether certain content is illegal.¹⁷⁶

A key anterior question in these cases will be who decides whether a take-down decision was taken due to local legal requirements or Facebook's own Community Standards. If Facebook itself decides this question, this too undermines the FOB's oversight capacity by giving Facebook a way of deciding that certain cases should not reach the FOB. Facebook could determine, without the possibility of review, that the FOB cannot hear a particular case. This may be exacerbated by the fact that when legal liability is not clear, platforms tend to err on the side of caution.¹⁷⁷

The second notable thing about the subject-matter jurisdiction of the FOB is its overall narrowness. Apart from possible abstract review referrals by Facebook, the FOB is limited to reviewing individual applications of Facebook's Community Standards. The biggest controversies around content moderation for Facebook in the last few years have concerned its decisions to take down, or not take down, putative hate speech,¹⁷⁸ foreign interference in domestic elections,¹⁷⁹ and other forms of misinformation and disinformation (including so-called "fake news,"¹⁸⁰ "deepfakes" and

[<https://perma.cc/G6QE-2ABU>] (summarizing laws targeting disinformation alone).

¹⁷⁶ See, e.g., Nele Achten, *Social Media Content Moderation: The German Regulation Debate*, LAWFARE (Dec. 27, 2018), <https://www.lawfareblog.com/social-media-content-moderation-german-regulation-debate> [<https://perma.cc/R8M3-3JYC>] (describing German rulings on NetzDG laws).

¹⁷⁷ DAPHNE KELLER, HOOVER INST., OBSERVATIONS ON SPEECH, DANGER, AND MONEY 2 (2018); Kaye, *supra* note 14, at 7.

¹⁷⁸ Issie Lapowsky, *Facebook Moves to Limit Toxic Content as Scandal Swirls*, WIRED (Nov. 15, 2018), <https://www.wired.com/story/facebook-limits-hate-speech-toxic-content/> [<https://perma.cc/5AEZ-6PJC>].

¹⁷⁹ Sheera Frenkel & Mike Isaac, *Facebook 'Better Prepared' to Fight Election Interference, Mark Zuckerberg Says*, N.Y. TIMES (Sept. 13, 2018), <https://www.nytimes.com/2018/09/13/technology/facebook-elections-mark-zuckerberg.html> [<https://perma.cc/YQ5C-6F8Q>].

¹⁸⁰ Craig Silverman, *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook*, BUZZFEED NEWS (Nov. 16, 2016), <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook> [<https://perma.cc/MDX3-5ATK>]; Adam Mosseri, *Working to Stop Misinformation and False News*, FACEBOOK MEDIA

“cheapfakes”¹⁸¹). But these are not the only important types of content moderation decisions Facebook makes. Allowing tech companies to frame “content decisions” or “content moderation” in such a limited way would get in the way of creating meaningful oversight and instead result in a type of transparency theatre.

As examples, there are at least two other key areas of “content decisions” that the FOB should have jurisdiction to adjudicate disputes under, both of which would be a more significant abdication of authority by the company. The first is algorithmic ranking decisions, and in particular the decision to down-rank certain content and decrease its circulation. The second is the application of its advertising policies, especially those around compliance with the additional new requirements Facebook has put in place concerning political ads.

Both of these decisions go much more directly to the core of Facebook’s business model than individual content moderation decisions about whether user posts comply with the Community Standards. The Facebook News Feed algorithm is Facebook’s “secret sauce,”¹⁸² that drives user engagement. Facebook is

(Apr. 7, 2017), <https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news> [<https://perma.cc/4UVB-6HB9>].

¹⁸¹ Robert Chesney et al., *About That Pelosi Video: What to Do About ‘Cheapfakes’ in 2020*, LAWFARE (May 29, 2019), <https://www.lawfareblog.com/about-pelosi-video-what-do-about-cheapfakes-2020> [<https://perma.cc/STC4-VCWS>].

¹⁸² THE FACEBOOK DILEMMA, Frontline, Part I at 12:22 (Antonio Garcia Martinez, former product manager at Facebook).

notoriously protective of its algorithm,¹⁸³ and users,¹⁸⁴ academics,¹⁸⁵ civil society,¹⁸⁶ and lawmakers¹⁸⁷ have long been calling for greater transparency around how Facebook determines what gets shown in the News Feed. Advertising is the backbone of Facebook's business.¹⁸⁸ Facebook might be concerned that giving the FOB power to review decisions in these domains risks too much intervention in the platform's core product design and revenue stream. But denying this jurisdiction undermines any claim that the

¹⁸³ Will Oremus, *Who Controls Your Facebook Feed*, SLATE (Jan. 3, 2016), http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html [<https://perma.cc/GX6A-6R7Y>]; Max Mason, *Facebook lashes algorithm regulatory oversight as "unworkable,"* AUSTRALIAN FIN. REV. (Dec. 12, 2018), <https://www.afr.com/business/media-and-marketing/advertising/facebook-lashes-algorithm-regulatory-oversight-as-unworkable-20181212-h191da> [<https://perma.cc/4WRC-QG26>].

¹⁸⁴ Pete Brown, *Study: Readers Are Hungry for News Feed Transparency*, COLUM. JOURNALISM REV. (Oct. 24, 2017), https://www.cjr.org/tow_center/study-readers-hungry-news-feed-transparency-algorithms.php [<https://perma.cc/YS2D-G8RF>].

¹⁸⁵ DAVID M.J. LAZER ET AL., ADDRESSING FAKE NEWS REQUIRES A MULTIDISCIPLINARY EFFORT 4 (2018).

¹⁸⁶ Corynne McSherry Cohn, Jillian C. York, & Cindy Cohn, *Private Censorship Is Not the Best Way to Fight Hate or Defend Democracy: Here Are Some Better Ideas*, ELECTRONIC FRONTIER FOUND. (Jan. 30, 2018), <https://www EFF.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some> [<https://perma.cc/LKL6-A6PC>]; Emma Llanso, *Is Holocaust Denial Free Speech? Facebook Needs to Be More Transparent*, FORTUNE (July 24, 2018) <http://fortune.com/2018/07/24/facebook-mark-zuckerberg-holocaust-denial-free-speech/> [<https://perma.cc/9GSC-AUKE>].

¹⁸⁷ Alexander Pirang, *Germany Wants Greater Algorithmic Transparency to Fight Disinformation, But Its Approach Is Half-Baked*, COUNCIL ON FOREIGN REL. (Apr. 11, 2018), <https://www.cfr.org/blog/germany-wants-greater-algorithmic-transparency-fight-disinformation-its-approach-half-baked> [<https://perma.cc/AD94-L52D>]; Evelyn Douek, *Congress' Grilling of Tech Companies in 2017 Foreshadows the Debates of 2018*, LAWFARE (Jan. 11, 2018), <https://www.lawfareblog.com/congress-grilling-tech-companies-2017-foreshadows-debates-2018> [<https://perma.cc/A4NH-7YXC>].

¹⁸⁸ Emily Stewart, *Mark Zuckerberg Testimony: Senators Seem Really Confused about Facebook*, VOX (Apr. 10, 2018), <https://www.vox.com/policy-and-politics/2018/4/10/17222062/mark-zuckerberg-testimony-graham-facebook-regulations> [<https://perma.cc/F4WC-RAVT>] (when asked by Senator Orrin Hatch about how Facebook sustained a business model when users do not pay for their service, Zuckerberg famously replied "Senator, we run ads.").

FOB is a meaningful and bona fide attempt to give greater rigor and transparency to Facebook's content moderation ecosystem as a whole. Indeed, confining the scope of jurisdiction of courts is a common technique used by authoritarian regimes to undermine the check on power that an independent judiciary might otherwise provide.¹⁸⁹ As Martin Shapiro has observed, "[a] relatively independent judiciary may be preserved but simply excluded from domains significant to the authoritarian regime."¹⁹⁰ As the rest of this section shows, both ranking and advertising decisions are core content moderation matters, and a body that is intended to provide "oversight" to content decisions needs to be empowered to review them.

A Degree of Algorithmic Transparency

Often, content moderation conversations revolve around a "take down / leave up" dichotomy. But platforms have far greater capacity to control the content on their sites than this paradigm suggests. Facebook is increasingly relying not on the blunter content moderation tools of removing posts or pages, but on the subtler tools of limiting their reach and exposure. For "borderline" content in each of its harmful categories, Facebook works to "distribute that content less" to reduce the incentive to post such content.¹⁹¹ Zuckerberg argues that "no matter where we draw the lines for what is allowed, as a piece of content gets close to that line, people will engage with it more on average."¹⁹² But Facebook's decision to "downrank" a piece of content (or distribute it less) in users' News

¹⁸⁹ Moustafa, *supra* note 36; Law, *supra* note 135, at 812 n.150 (describing how Taiwan's legislative body changed the quorum requirements for the issuance of constitutional interpretations in order to substantially diminish such rulings); CARLO GUARNIERI & PATRIZIA PEDERZOLI, *THE POWER OF JUDGES: A COMPARATIVE STUDY OF COURTS AND DEMOCRACY* 78–79 (Cheryl Thomas trans., Oxford Socio-Legal Studies, 2002) ("[T]he kind of disputes a judge can be asked to settle . . . provides the starting point for assessing the scope of courts. . . . Authoritarian regimes often adopt [a] fragmentation strategy to control the administration of justice."); Landau & Dixon, *supra* note 81, at 30.

¹⁹⁰ Martin Shapiro, *Courts in Authoritarian Regimes*, in *RULE BY LAW: THE POLITICS OF COURTS IN AUTHORITARIAN REGIMES* 326, 331 (Tamir Moustafa & Tom Ginsburg eds., 2008).

¹⁹¹ Zuckerberg, *supra* note 1.

¹⁹² *Id.*

Feeds is currently much less transparent than a decision to take down a piece of content. Users are typically notified when a post is removed entirely, but, because users are not told how their posts are treated by the News Feed algorithm, may be entirely unaware when their post is left up but just not shown to other users. From Facebook's perspective, this is a key virtue of this approach. As the company notes in a recent patent for this practice, when it is used "the commenting user is not made aware that his or her comment was blocked, thereby providing fewer incentives to the commenting user to spam the page or attempt to circumvent the social networking system filters."¹⁹³

Opacity in Facebook's ranking decisions leaves users guessing at what is happening, and speculating if this is the cause of any drop in engagement with their posts. Some (including President Trump) have suggested that social media platforms "shadow ban"¹⁹⁴ posts simply because they do not like them.¹⁹⁵

Crucially, if the FOB cannot review down ranking decisions not only will this speculation persist, but also Facebook can still effectively control the extent to which the FOB can provide oversight: if Facebook does not want a takedown to be subject of public attention and the FOB's review, Facebook could simply downrank that content effectively out of circulation instead of removing it.

Facebook's content moderation includes decisions about how to distribute content Facebook does *not* take down. Therefore, to give legitimacy to this wider ecosystem the FOB needs jurisdiction over these decisions. Of course, in a sense, every piece of content on Facebook is subject to a degree of algorithmic content moderation.

¹⁹³ Bryan Menegus, *Facebook Patents Shadowbanning*, GIZMODO (July 16, 2019), <https://gizmodo.com/facebook-patents-shadowbanning-1836411346> [<https://perma.cc/YTD8-N4HP>].

¹⁹⁴ G.F., *What is "Shadowbanning"?*, ECONOMIST (Aug. 1, 2018), <https://www.economist.com/the-economist-explains/2018/08/01/what-is-shadowbanning> [<https://perma.cc/L8RM-GCQN>].

¹⁹⁵ Tal Axelrod, *Trump: Facebook, Twitter and Google Bias is 'Ridiculous.'* HILL (Dec. 18, 2018), <https://thehill.com/homenews/administration/421820-trump-facebook-twitter-and-google-bias-is-ridiculous> [<https://perma.cc/K2W5-2QUQ>].

Facebook has famously jealously guarded the workings of its News Feed algorithm, to prevent user gaming and for trade secrets reasons. There may be a legitimate argument that the FOB cannot or should not have complete oversight of all aspects of the algorithm; but at least to the extent that algorithmic changes are used explicitly as a tool of content moderation on the basis of lines drawn in the Community Standards, these should be subject to FOB review.

Advertising Policies

Facebook has detailed policies on advertising and all ads purchased undergo pre-publication review to ensure they meet Facebook's rules.¹⁹⁶ Facebook has made significant changes to its political advertising policies in the wake of controversy surrounding the 2016 U.S. election. Advertisers that run political ads are now required to go through an authorization process and run disclaimers about who paid for the ads.¹⁹⁷ Facebook's definition of a political ad includes an ad that relates to any of twenty "issues of national importance," a list that includes, for example, the indeterminate category of "values."¹⁹⁸ Facebook's application of these rules has given rise to controversy. Ads from news sites¹⁹⁹ and non-profits²⁰⁰ were removed because they were flagged as "political" but had not

¹⁹⁶ *Advertising policies*, FACEBOOK, <https://www.facebook.com/policies/ads#> [<https://perma.cc/8WZV-GX5Z>].

¹⁹⁷ *Id.* ("5. Restricted Content . . . 11.a. Ads related to politics or issues of national importance").

¹⁹⁸ *Issues of national importance*, FACEBOOK ADS HELP CENTRE, <https://en-gb.facebook.com/business/help/214754279118974> [<https://perma.cc/25BY-EJN2>].

¹⁹⁹ Ariana Tobin & Jeremy B. Merrill, *Facebook's Screening for Political Ads Nabs News Sites Instead of Politicians*, PROPUBLICA (June 15, 2018), <https://www.propublica.org/article/facebook-new-screening-system-flags-the-wrong-ads-as-political> [<https://perma.cc/S63B-98QK>].

²⁰⁰ Daniel Gallant, *Facebook Censors at Random*, WALL ST. J. (Dec. 9, 2018), <https://www.wsj.com/articles/facebook-censors-at-random-1544395970> [<https://perma.cc/HY5B-DXG4>]; J. Nathan Matias, Austin Hounsel, & Melissa Hopkins, *We Tested Facebook's Ad Screeners and Some Were Too Strict*, ATLANTIC (Nov. 2, 2018), <https://www.theatlantic.com/technology/archive/2018/11/do-big-social-media-platforms-have-effective-ad-policies/574609/> [<https://perma.cc/H3GV-G5SV>].

received the necessary special authorization, while ads purchased in the names of sitting Senators were waved through.²⁰¹ The individual and societal consequences of inconsistent and opaque enforcement of political advertising policies are potentially significant. Yet, as Kreiss and McGregor found, Facebook “exercise[s] considerable discretion over paid political media. While these decisions are clearly difficult, they are made without much in the way of transparency, consistency, or accountability.”²⁰² The lack of explanation for how these policies are applied is particularly concerning because without it “Facebook would have the authority to determine the messages that campaigns pay for the public to see, without the [public] challenges that keep these firms learning and cause their policies to evolve.”²⁰³ Without transparency, it also opens Facebook up to the charge that its decisions are politically motivated.

Inconsistent and unexplained decision-making is exactly the problem the FOB is being created to solve. In the context of political content, this is especially important. The mere fact that content is paid and not “organic” does not alleviate the need for oversight. Giving the FOB’s jurisdiction in these cases would put literal money where Facebook’s mouth is about trusting the expertise of the FOB to interpret and apply its rules in a legitimate manner.

In the report summarizing the results of its global consultation on the FOB, Facebook said that it had been “clear” that the Board was only going to review individual decisions under the Community Standards and that while News Feed ranking and political advertising were “important issues,” they were beyond the remit of the Board.²⁰⁴ But these are not just important issues: they are at the core of Facebook’s content moderation ecosystem. Excluding them

²⁰¹ William Turton, *We Posed as 100 Senators to Run Ads on Facebook. Facebook Approved All of Them.*, VICE NEWS (Oct. 30, 2018), https://news.vice.com/en_ca/article/xw9n3q/we-posed-as-100-senators-to-run-ads-on-facebook-facebook-approved-all-of-them [<https://perma.cc/P37T-9GTU>].

²⁰² Daniel Kreiss & Shannon C. McGregor, *The “Arbiters of What Our Voters See:” Facebook and Google’s Struggle with Policy, Process, and Enforcement Around Political Advertising*, POLITICAL COMM. 1, 2 (2019).

²⁰³ *Id.* at 11–12.

²⁰⁴ Harris, *supra* note 138.

from the FOB's jurisdiction undermines Facebook's broader commitment to transparency and accountability in content moderation that the FOB is intended to facilitate.

IV. THE LEGITIMACY CONUNDRUM

The fundamental purpose of the FOB is to bring greater legitimacy to Facebook's content moderation ecosystem. While enhancing the legitimacy of the FOB *itself* is instrumental to this goal, it is not the ultimate aim: the aim is to legitimize *Facebook's* exercise of power by subjecting it to an independent check. The focus of this article is therefore a more general sociological legitimacy—that is, the extent to which the public regards Facebook's decisions generally as justified, appropriate, or otherwise deserving of support beyond the fear of sanctions or mere hope for personal reward.²⁰⁵

In a crucial sense, this legitimacy can only be established over time. For the FOB itself, as the Draft Charter notes, “[t]he public legitimacy of the board will grow from the transparent, independent decisions that the board makes.”²⁰⁶ But gaining legitimacy will be difficult. The FOB is being established in response to widespread public dissatisfaction with Facebook content moderation and so the enterprise begins on the back foot. And while the early use of the moniker of a “Supreme Court” may have been (intentionally or unintentionally) intended to invoke the aura of legitimacy that the Supreme Court of the United States enjoys, the FOB will not have a reservoir of legitimacy accumulated over a long history to draw down upon.²⁰⁷

But there are two more fundamental challenges in the quest for legitimacy: first, the difficulty for Facebook making a credible commitment to being bound by the FOB when, ultimately, Facebook itself retains final authority; and second, the lack of agreed prior norms or authorities for content moderation which the

²⁰⁵ Richard H. Fallon, *Legitimacy and the Constitution*, 118 HARV. L. REV. 1787, 1795 (2004–2005) (citing TOM R. TYLER, WHY PEOPLE OBEY THE LAW 26 (1990)).

²⁰⁶ DRAFT CHARTER, *supra* note 2, at 5.

²⁰⁷ See Fallon, *supra* note 205, at 1829, 1837–38.

FOB can base its decisions on and leverage into accepted pronouncements. This part addresses each of these challenges in turn.

A. The Limits of the FOB's Legitimizing Power

There is a real limit on the extent of the legitimation that the FOB can bring to Facebook content moderation. Part II(B) above discussed the ways in which authoritarian regimes use courts to give them a veneer of legitimacy. But as Mark Tushnet has pointed out about this literature, “[t]he general difficulty with these accounts is straightforward: rulers might want to make credible commitments, but they cannot do so, precisely because they can alter the constitution whenever they want—and the target audiences know that these rulers can do so.”²⁰⁸ In Facebook’s case, the conundrum is the same: Zuckerberg wants to assure stakeholders that the FOB will be independent, make decisions “in the best interests of our community and not for commercial reasons,”²⁰⁹ and that Facebook will consider itself bound by the FOB’s decisions. But it is difficult for Facebook to make this commitment to tie itself to the mast of the FOB completely credible. Ultimately, Facebook does not have to obey the FOB’s decisions or could disband the institution altogether without cause. This is the flip side of judicial review being a relatively low-cost way of enhancing legitimacy without opening up the broader policy-making process—you get what you pay for.

A practical issue may also arise for Facebook of how to resolve its fiduciary duties to its shareholders should the FOB issue a decision that it genuinely regarded as against its commercial interests.²¹⁰ The simple response is that the FOB can be considered in the long-term commercial interests of the company because it increases user satisfaction, protects company reputation, and potentially staves off more severe regulation. But to the extent that this is true, the FOB does not actually constitute a “check” on Facebook’s power because its actions remain in Facebook’s best

²⁰⁸ Tushnet, *supra* note 92, at 422–23.

²⁰⁹ Zuckerberg, *supra* note 1.

²¹⁰ Lina Khan & David Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. (forthcoming 2019).

interests. As Tushnet explains, when it serves the interest of the authoritarian for there to be an external check on its powers:

[C]ourts do no work because the regime's immediate self-interest will lead it to refrain from actions that reduce the returns it anticipates to gain during the period when the preferences are stable. And, if the regime's preferences change, the mechanisms also do no work because the regime is free to change them to accommodate its new preferences.²¹¹

Similarly, should Facebook decide that the cost of public legitimacy becomes too high, it can simply abandon its experiment.²¹²

The notion that the FOB is merely instrumental to Facebook and does not constitute a meaningful check is further suggested by a number of design choices. The highly circumscribed and subjective nature of the FOB's subject-matter jurisdiction, discussed above,²¹³ is the most significant. But there is also the fact that Facebook has said that the company “*can* incorporate the board's decisions in the policy development process,”²¹⁴ suggesting that this possibility is discretionary. There is a fundamental tension for Facebook in creating a meaningfully independent body while also keeping its discretion within acceptable bounds. As Zuckerberg asked, “[h]ow do we ensure their independence from Facebook, but also their commitment to the principles they *must* uphold?”²¹⁵ Facebook's answer to this has been to require FOB members to agree to uphold a set of values specified by Facebook.²¹⁶

²¹¹ Tushnet, *supra* note 92, at 425; *see also* Khan & Pozen, *supra* note 210, at 10.

²¹² For discussion of the fact that the FOB represents a “bet” that public legitimacy is a worthwhile commercial investment that could be abandoned, *see* Evelyn Douek, *YouTube's Bad Week and the Limitations of Laboratories of Online Governance*, LAWFARE (June 11, 2019), <https://www.lawfareblog.com/youtubes-bad-week-and-limitations-laboratories-online-governance> [<https://perma.cc/9PZG-8JZY>].

²¹³ *See infra* Part III(C).

²¹⁴ DRAFT CHARTER, *supra* note 2, at 1 (emphasis added).

²¹⁵ Zuckerberg, *supra* note 1 (emphasis added).

²¹⁶ DRAFT CHARTER, *supra* note 2, at 5.

B. The Limits of a “Constitution’s” Legitimizing Power

Requiring the FOB’s decisions to be based on a set of underlying values has been likened to the adoption of a “constitution” which will guide the FOB’s interpretation of the Community Standards.²¹⁷ These values will play an important role in the FOB’s work, but there are crucial ways in which they are unlike a constitution. A values statement can perform the same role as constitutions do in expressing Facebook’s fundamental vision for its platform which the FOB can use to resolve ambiguity in the Community Standards in a way that reflects these commitments. But because the values will be Facebook’s and not the user community’s, they cannot perform the same role of legitimization that constitutions do because there is no sense in which the values express a delegation of authority from users or a set of widely-agreed norms.

The statement of values will play an essential role by guiding the FOB in difficult cases where there is no clear answer. It is precisely because there will be instances where competing understandings of the Community Standards are possible that the FOB is necessary. If the Community Standards provided a clear answer in every case, there would be no need to seek review of content moderation decisions. In cases where there are multiple possible interpretations of these rules, what should the FOB use to decide between them? Without a set of underlying values, decisions become relatively unconstrained. This is suboptimal from Facebook’s perspective, because it creates a large sphere of discretion for FOB members. But it is also suboptimal for users and for the legitimacy of the FOB itself: without an underlying set of commitments, decisions can become unpredictable or seen as arbitrary.

In its Draft Charter, Facebook said it would adopt a set of values that “would encompass concepts like voice, safety, equity, dignity, equality and privacy.”²¹⁸ Importantly, a list of vague “values” does not help achieve the aim of guiding the FOB’s interpretative practice. A list that includes everything prioritizes nothing. It is hard to disagree with the importance of each of “voice, safety, equity, dignity, equality and privacy,” but the very nature of hard freedom

²¹⁷ Klonick & Kadri, *supra* note 128.

²¹⁸ DRAFT CHARTER, *supra* note 2, at 5.

of speech cases is that they involve trade-offs between these values. The difference between various countries' hate speech jurisprudence, for example, is largely due to a different weighting of these values. As Rosenfeld explains:

If free speech in the United States is shaped above all by individualism and libertarianism, collective concerns and other values such as honor and dignity lie at the heart of the conceptions of free speech that originate in international covenants or in the constitutional jurisprudence of other Western democracies. Thus, for example, Canadian constitutional jurisprudence is more concerned with multiculturalism and group-regarding equality. For its part, the German Constitution sets the inviolability of human dignity as its paramount value. . . . These differences have had a profound impact on the treatment of hate speech.²¹⁹

Albie Sachs has similarly observed:

To Americans, the firstness of the First Amendment is axiomatic. It is seen as a source of enlightenment, as being the most constitutive and defining element of the whole constitutional order. The legal cultures of Germany and South Africa, however, have a profoundly different foundational element. It is not free speech, but human dignity. What is axiomatic to an American lawyer could be problematic to us. What is axiomatic to us could be problematic to an American.²²⁰

Just as AI cannot perform content moderation without coders making value judgments in writing the algorithms,²²¹ so too Facebook cannot escape the need to make a choice about the kind of platform it wants to be. The decision of whether to allow conspiracy-monger Alex Jones to continue to stream and share violent and sometimes racist conspiracy theories is an example where such a choice is necessary.²²² On one side is the value of voice and liberty

²¹⁹ Michel Rosenfeld, *Hate Speech in Constitutional Jurisprudence: A Comparative Analysis Conference: The Inaugural Conference of the Floersheimer Center for Constitutional Democracy: Fundamentalisms, Equalities, and the Challenge to Tolerance in a Post-9/11 Environment*, 24 CARDOZO L. REV. 1523, 1541 (2002–2003).

²²⁰ Albie Sachs, *Reflections on the Firstness of the First Amendment in the United States*, in *THE FREE SPEECH CENTURY* 179, 192 (Lee C. Bollinger & Geoffrey R. Stone eds., 2018).

²²¹ See *supra* Part II(A).

²²² Jane Coaston, *YouTube, Facebook, and Apple's Ban on Alex Jones, Explained*, VOX (Aug. 6, 2018), <https://www.vox.com/2018/8/6/17655658/alex-jones-facebook-youtube-conspiracy-theories> [<https://perma.cc/N9AF-HAVE>];

to speak, but on the other hand there is the safety and dignity of people affected by his hateful rhetoric or conspiracy-mongering. Another example is the decision to prohibit content promoting eating disorders, which reflects a prioritization of safety over unconstrained voice.²²³ These choices are controversial—this is part of the reason the FOB is being set up. But an uncontroversial statement of values that is universally acceptable would be stated at such a level of generality to be of limited utility in constraining decision-making.

Zuckerberg has in fact long articulated underlying values he sees as inherent in his platform. In the opening words of his Blueprint announcing the FOB's creation he says “[m]any of us got into technology because we believe it can be a democratizing force for putting power in people’s hands. I’ve always cared about this and that’s why the first words of our mission have always been “give people the power.”²²⁴ This kind of rhetoric attracts backlash, but a transparent statement of what the platform is intended to be is important in explaining choices made in content moderation. And by explicitly choosing priorities in advance, this reduces the chances that the FOB’s decision-making becomes so out of step with the company’s vision for its platform that it decides to simply ignore or overrule its decisions.

But although a charter that makes difficult choices will be necessary for the FOB to function effectively, it cannot cure Facebook’s legitimacy deficits for several reasons. First, as discussed, there is no mechanism to double-entrench any “constitution” such that Facebook could not amend it at-will, making it difficult to place too much faith in its power. Second, the constitution itself will not have sociological legitimacy and so users

Kurt Wagner, *Facebook is Taking Down 22 More Pages Tied to Infowars Founder Alex Jones*, RECODE (Feb. 5, 2019), <https://www.recode.net/2019/2/5/18212439/facebook-alex-jones-remove-pages-infowars-again> [<https://perma.cc/393M-XXSQ>].

²²³ Laurence Dodds, *Facebook and Instagram Crack Down on Anorexia Posts in Effort to Combat Eating Disorders*, TELEGRAPH (July 13, 2019), <https://www.telegraph.co.uk/technology/2019/07/13/facebook-instagram-ban-anorexia-posts-effort-combat-eating-disorders/> [<https://perma.cc/W53T-SRMV>].

²²⁴ Zuckerberg, *supra* note 1.

have no special reason to accept decisions made ostensibly in accordance with it. A list of human rights buzzwords released by Facebook or chosen by Zuckerberg will not have the same resonance as national constitutions that are typically adopted through democratic processes. Although Zuckerberg seems to have been reading some Rawls,²²⁵ he is more founder than founding father and has little mandate to decide the values of the entire community of Facebook users. A choice by him to prioritize “voice” over “dignity,” for example, might resonate with some users but cannot be said to be reflective of the will of those users in any real sense nor will those that disagree have much reason to feel themselves reasonably beholden to Zuckerberg’s choice.

Facebook could, and many have argued should,²²⁶ adopt other widely endorsed norms such as international human rights law as the basis of its decisions. Facebook itself has said it looks to documents like the International Covenant on Civil and Political Rights for guidance on where to draw the lines on freedom of expression.²²⁷ There are some potential limitations with adopting these norms wholesale,²²⁸ but they could provide a starting point on which to develop a more mature form of content moderation. International human rights law has the advantage of having a significant reservoir of legitimacy in certain communities which Facebook and the FOB could draw down on. Many scholars are doing important work on this topic, but it is not explored any further

²²⁵ John Constine, *Highlights & Transcript from Zuckerberg’s 20K-word Ethics Talk*, TECHCRUNCH (Feb. 20, 2019), <https://techcrunch.com/2019/02/20/zuckerberg-harvard-zittrain/> [<https://perma.cc/YJ94-NKZH>].

²²⁶ Kaye, *supra* note 14; Evelyn Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26 (2018); ASH ET AL., *supra* note 16, at 7; see discussion of this proposal in Evelyn Douek, *U.N. Special Rapporteur’s Latest Report on Online Content Regulation Calls for “Human Rights by Default,”* LAWFARE (June 6, 2018), <https://www.lawfareblog.com/un-special-rapporteurs-latest-report-online-content-regulation-calls-human-rights-default> [<https://perma.cc/3YXC-JF4Y>].

²²⁷ Richard Allan, *Hard Questions: Where Do We Draw the Line on Free Expression?*, FACEBOOK NEWSROOM (Aug. 9, 2018), <https://newsroom.fb.com/news/2018/08/hard-questions-free-expression/> [<https://perma.cc/DGR9-UEMY>].

²²⁸ This is discussed in Douek, *supra* note 226. I intend to explore these issues further in later work.

here for the simple reason that at this stage Facebook is showing no indication that it intends for international human rights law to be the fundamental basis for its content moderation.

Therefore, a fundamental set of values is a necessary but not sufficient part of creating legitimacy for Facebook's content moderation and the FOB's decisions. Affirmatively stating what the purpose of content moderation on Facebook is can guide the FOB's interpretative process, but it cannot of itself compel those who had no say in deciding those values to view them as legitimate.

V. MAXIMIZING THE FOB'S POTENTIAL

The FOB can still bring significant benefits to Facebook's content moderation despite these limits on the legitimacy it can create. Maximizing these benefits requires acknowledging these limitations and designing around them. This includes accepting two facts inherent in the problem of content moderation on private platforms: Facebook will always retain the power to overrule the FOB's decisions, and in many of the most difficult cases there will be no "right" answer. Institutional design that accounts for these facts will be stronger in the long run. It should focus on the two key benefits that the FOB can bring. First, a practical benefit: a judicial-style check, even if reversible by Facebook, can improve Facebook's policy formation processes. Second, a sociological benefit: bringing greater public acceptance of Facebook's rules through the performance of public reasoning.²²⁹

A. *Weakness as Strength*

The FOB can improve Facebook's policies without being the ultimate authority on every aspect of the platform's rules. The idea that the FOB's review of Facebook's decisions should be final and irreversible, not only in the particular case but in all similar cases going forward, is a fundamentally American perspective. Indeed, this is known as strong-form judicial review, and the Supreme Court of the United States is considered the "archetype or paradigm" of

²²⁹ This Part expands on some of the arguments made in Douek, *supra* note 124.

this practice.²³⁰ Tushnet, the leading writer on the strong-form/weak-form distinction, describes the strong-form model in the following way: “[i]n systems with strong-form judicial review, of which the United States is usually taken to be representative, the constitutional court has the power to invalidate primary legislation, and legislative responses to such invalidations are made quite difficult.”²³¹

By contrast, the central characteristic of weak-form review is that “courts assess legislation against constitutional norms, but do not have the final word on whether statutes comply with those norms” and judicial interpretation can be displaced by ordinary legislative processes.²³² Commentators have suggested that this kind of override would undermine the FOB’s independence. Kadri has argued that “[i]f Facebook is free to unilaterally overrule appellate decisions it doesn’t like, the talk of a new era of radical transparency and accountability will be overblown. . . . [A]mending its legal code willy-nilly would undermine the entire project.”²³³ But the comparative literature on “weak-form” or “dialogic” review suggests this may be too simplistic. There are many benefits of a weak-form model of review that may make it more appropriate for the dynamic and complex environment that is Facebook.

1. *The Benefits of a Judicial-Style Check on Policy-Making*

Weak-form review is well-suited to counter blockages in the “legislative process” (here, the formulation of Community Standards).²³⁴ Dixon describes two forms of blockages: blind spots and inertia. Both are present in Facebook’s policy formation.

Blind spots arise because initial policies are often written in time-pressured conditions, and cannot fully anticipate all the

²³⁰ Aileen Kavanagh, *What’s so Weak about “Weak-Form Review”? The Case of the UK Human Rights Act 1998*, 13 INT’L J. CONST. L. 1008, 1011 (2015).

²³¹ Mark Tushnet, *Judicial Activism or Restraint in a Section 33 World Review Article*, 53 U. TORONTO L.J. 89, 89 (2003).

²³² MARK TUSHNET, *WEAK COURTS, STRONG RIGHTS: JUDICIAL REVIEW AND SOCIAL WELFARE RIGHTS IN COMPARATIVE CONSTITUTIONAL LAW* ix (2008).

²³³ Thomas Kadri, *How Supreme a Court?*, SLATE MAGAZINE (Nov. 19, 2018), <https://slate.com/technology/2018/11/facebook-zuckerberg-independent-speech-content-appeals-court.html> [<https://perma.cc/AA9X-X3RR>].

²³⁴ Dixon, *supra* note 27.

possible circumstances in which a policy will need to be applied.²³⁵ This well describes the process for the formulation of Community Standards at Facebook. Initial rules were written haphazardly as international expansion meant the platform needed to accommodate a rapidly growing and changing user base, and revisions have often been prompted by particular high-profile controversies.²³⁶ As Klonick has observed, “internal policies and the rules that reflect them are constantly being updated . . . because Facebook is attempting, in large part, to rapidly reflect the norms and expectations of its users.”²³⁷ A recent Oxford-Stanford report highlighted that “Facebook’s policy changes and priorities are often highly reactive to outrage and scandal in the relatively narrow spectrum of Western media and politics.”²³⁸ This description of the reactive nature of Facebook’s policy-making parallels the description given by Calabresi of the typical legislative process, who notes that blind spots can arise because “[l]egislatures often act hastily or thoughtlessly with respect to fundamental rights because of panic or crises or because, more often, they are simply pressed for time.”²³⁹ Judicial-style checks help overcome blind spots, because they consider the application of rules retrospectively and in the context of particular cases.²⁴⁰ The FOB can take into account perspectives that initial policy-makers did not have available to them or provide an opportunity to review unintended consequences of a rule’s operation in practice.

²³⁵ *Id.* at 2208–09.

²³⁶ Klonick, *supra* note 40, at 1630–35, 1648–58 (discussing updates to policies as a result of public outcry regarding Facebook’s treatment of posts depicting breastfeeding, a gay kiss, nude art, beheading videos and a number of other examples.); GILLESPIE, *supra* note 29, at 67 (“Revisions of the guidelines often come only in response to outcries and public controversies.” Gillespie discusses the example of pro-ana and self-harm guidelines updated in response to a media expose.).

²³⁷ Klonick, *supra* note 40, at 1649.

²³⁸ ASH ET AL., *supra* note 16, at 9.

²³⁹ Guido Calabresi, *Foreword: Antidiscrimination and Constitutional Accountability (What the Bork-Brennan Debate Ignores)*, 105 HARV. L. REV. 80, 103–4 (1991).

²⁴⁰ Dixon, *supra* note 27, at 2214; Richard H. Fallon, *The Core of an Uneasy Case for Judicial Review*, 121 HARV. L. REV. 1693, 1697 (2008).

Inertia arises where policy-makers do not have the time to devote to changing rules that are already in effect or because the status quo will often prevail in hard cases when there are disagreements about what should be done.²⁴¹ This can be seen in the history of content moderation on Facebook too. For example, the status quo of leaving up Alex Jones' posts prevailed until the company was forced to reconsider because other platforms decided to remove him.²⁴² The FOB can create a more formalized process for prompting Facebook to address blockages. The process of review can disrupt the status quo, draw public and Facebook's attention to blind spots or inertia, and provide an opportunity for Facebook to correct any error.²⁴³

This form of check also capitalizes on various actors' specializations.²⁴⁴ FOB members will have expertise in "content, privacy, free expression, human rights, journalism, civil rights, safety and other relevant disciplines."²⁴⁵ But Facebook staff will have better understanding of the overall dynamics of speech on Facebook and the surrounding platform architecture. A FOB empowered with a weak-form judicial review mandate can create an ecosystem that puts these two kinds of expertise in "dialogue."

The FOB does not need to have ultimate authority over Facebook's rules to bring these benefits. First, even without such authority, Facebook overruling the FOB is likely to be rare. In practice, there would be high reputational costs for Facebook in disregarding a decision of the FOB. Because Facebook has been publicizing the FOB and extolling its benefits, there would likely be significant attention given to any substantial undermining of its authority, so Facebook is unlikely to take the decision lightly. If done often, it would give rise to the impression that the FOB is mere window-dressing and deprive it of all legitimacy. Furthermore, it is

²⁴¹ Dixon, *supra* note 27, at 2209–12.

²⁴² Kevin Roose, *Facebook Banned Infowars. Now What?*, N.Y. TIMES (Aug. 10, 2018), <https://www.nytimes.com/2018/08/10/technology/facebook-banned-infowars-now-what.html> [<https://perma.cc/4VZR-E2PY>].

²⁴³ Dixon, *supra* note 27, at 2216–19.

²⁴⁴ Mark Tushnet, *The Rise of Weak-Form Judicial Review*, in COMPARATIVE CONSTITUTIONAL LAW 325 (Tom Ginsburg & Rosalind Dixon eds., 2011).

²⁴⁵ DRAFT CHARTER, *supra* note 2, at 1.

important to remember that a key benefit of the FOB from Facebook's perspective is not the substantive rulings the FOB issues but the ability this gives Facebook to distance itself from controversial decisions. This incentive remains even where Facebook *could* overturn any decisions. The similar desire to blame courts for unpopular decisions means that in systems with weak-form review (such as the U.K.) political actors often use rhetoric that suggests they consider themselves bound by court decisions in order to pass the political buck.²⁴⁶ Therefore the practical strength of weak-form review is often much stronger than it appears in theory. As Feldman wrote in his original argument for a FOB, "As a corporation, [Facebook] has the right and capacity to change its policy at any time or even ignore it. The public understands this, as it should. . . . [But] [i]f Facebook were to violate its commitment, it would be subject to public censure and criticism."²⁴⁷

Additionally, in cases where Facebook does decide to overrule the FOB, it is likely to issue an explanation given the public attention that would surround the move. This reason-giving is also valuable, and not only as an improvement over the current (lack of) transparency that attends many changes in Facebook's policies. Explaining why Facebook took a different path to the one recommended by the FOB could improve the quality of debate around Facebook's rules more generally. Currently, contentious content decisions are made in ways that do not facilitate productive discussion, such as executives deciding matters in the early hours of the morning based on media reports of what other tech companies are doing.²⁴⁸ Introducing a judicial-style body into a system can elevate the mode of decision-making and discourse, "engender[ing] new modes of legislative discourse and practice."²⁴⁹ Mattias Kumm

²⁴⁶ Kavanagh, *supra* note 230, at 1022, 1027.

²⁴⁷ Harris, *supra* note 138, at 140.

²⁴⁸ Roose, *supra* note 242.

²⁴⁹ STONE SWEET, *supra* note 8, at 62, 74–75; Juliane Kokott & Martin Kaspar, *Ensuring Constitutional Efficacy*, in THE OXFORD HANDBOOK OF COMPARATIVE CONSTITUTIONAL LAW 806 (2012).

calls this “Socratic Constestation.”²⁵⁰ By putting decision-makers in dialogue with a judicial-style check, they begin to think about how their decisions can be justified in terms that are more publicly acceptable.

2. *The Advantages of Weak-Form Review*

The last section addressed the reasons why weak-form review will in most cases be sufficient to bring about the benefits of an oversight body. But there are also reasons why this model might in fact be preferable to strong-form review. This section addresses five reasons: two theoretical, and three pragmatic. Although less august, the pragmatic considerations may ultimately be more important.

First, weak-form review is more appropriate in cases that involve competing rights claims. Some proponents for judicial review acknowledge that strong-form review might be less appropriate when the judicial body will frequently have to decide zero-sum controversies involving the collision of two fundamental rights.²⁵¹ This is because such cases involve difficult trade-offs that are better left to a branch of government that is more democratically accountable or responsive. The speech disputes that the FOB will be charged with deciding, such as whether hate speech or sexually explicit materials should be taken down, are “common and readily expressible as ‘zero sum’ situations.”²⁵² These are archetypal clashes between liberty rights (to speak) and dignity or equality rights (for example, to be free from harassment). While Facebook itself is not democratically accountable like a legislature, it may still be more responsive to public pressure than a purposefully insulated FOB. Importantly, Facebook is also able to take proactive action in response to such pressure in the event that rules or FOB rulings become out of step with community values, rather than the FOB

²⁵⁰ Mattias Kumm, *The Idea of Socratic Contestation and the Right to Justification: The Point of Rights-Based Proportionality Review*, 4 L. & ETHICS OF HUM. RTS. 141 (2010).

²⁵¹ Fallon, *supra* note 240, at 1731; Mark Tushnet, *How Different Are Waldron’s and Fallon’s Core Cases For and Against Judicial Review?*, 30 OXFORD J. LEGAL STUD. 49, 55 (2010).

²⁵² Tushnet, *supra* note 251, at 55.

which will only make rulings in response to cases brought to it and not on its own initiative.

Second, because these cases involve such difficult trade-offs, where ultimately one person's right or interest must give way to another's, it is important to keep in mind the fragile legitimacy of the FOB itself. In such cases, the lack of a reservoir of goodwill on which to draw on to bolster decisions that are controversial and will upset large segments of the user base will be particularly noticeable. This is especially so given, as discussed above, the FOB's inability to be truly or comprehensively representative.²⁵³ This means that it will necessarily be making decisions to trade-off the rights or interests of people who will not feel they are represented in the decision-making process. While this can be somewhat mitigated by creating processes through which these communities are given a voice in proceedings, it will never be wholly resolvable. As such, it is better to be humble and not overstate the capacity of the FOB to be the final arbiter of norms for all of Facebook's diverse community and allow for feedback and development.

Third, and this is the critical pragmatic point, if the FOB's decisions are not seen as final and irrevocable, this may make Facebook more willing to give the FOB the broader jurisdiction that, as argued above, is necessary for it to have meaningful oversight of Facebook's content moderation as a whole. A narrowly confined and manipulatable jurisdiction will significantly impair the FOB's value. This is perhaps the single biggest threat to the FOB's legitimacy because it creates the impression that Facebook is only willing to renounce power in areas that do not "really matter" to it. However, it is also understandable that Facebook may not want to completely renounce power over this central aspect of its platform. Facebook's content moderation *is* in many ways its distinctive value offering.²⁵⁴ Facebook's right to determine the content that appears on its platform may also be expressed in terms of the "free speech" rights of platforms themselves.²⁵⁵ Weak-form review may be a

²⁵³ See Part III(A).

²⁵⁴ GILLESPIE, *supra* note 29, at 13.

²⁵⁵ See Heather Whitney, *Search Engines, Social Media, and the Editorial Analogy*, in KNIGHT FIRST AMENDMENT INSTITUTE EMERGING THREATS SERIES 8 (2018).

concession to this consideration, but also to encourage the benefits of dialogue between Facebook, the FOB, and stakeholders over a broader range of areas. This includes the FOB's ability to influence policy and not just rule on specific take-downs. This kind of abstract review jurisdiction is common in countries with specialized constitutional courts.²⁵⁶ This is because these courts were created, "explicitly and as a constitutional priority, to protect rights" and limiting the avenues to access judicial review would produce gaps in rights protection.²⁵⁷ And even in the U.S. there is a partial exception to the concrete case or controversy requirement for First Amendment challenges brought against laws for being vague or overbroad, given the potential chilling effect of such legislation.²⁵⁸ This is a recognition of not making the ability to bring challenges to speech rules too narrow or difficult. Therefore, a generally broad jurisdiction, including the ability to bring abstract review cases, is critical. Facebook is more likely to create this jurisdiction and make use of it more often if it retains the option to overrule decisions, this is a reasonable compromise. As Stone Sweet says of the constitutional courts of Europe that are more politicized but also more actively engaged in the protection of rights, "[t]he erosion of traditional separation of powers notions is the tax we pay for these benefits [of greater rights protection]."²⁵⁹

Fourth, a concession to realism: should the FOB make a ruling that is fundamentally damaging to Facebook or out-of-step with user values, Facebook is likely to step in and overrule it. Indeed, it may even have a duty to its shareholders to do so.²⁶⁰ It is better to accept this reality from the outset, instead of creating a system that will only be undermined.²⁶¹ This pragmatism will create a more robust and durable institution in the long run.

²⁵⁶ See Jenny S. Martinez, *Horizontal Structuring*, in THE OXFORD HANDBOOK OF COMPARATIVE CONSTITUTIONAL LAW 548, 572 (Michel Rosenfeld & Andr  s Saj   eds., 2012).

²⁵⁷ Stone Sweet, *supra* note 90, at 824.

²⁵⁸ *Bd. of Airport Comm'rs v. Jews for Jesus*, 482 U.S. 569, 576 (1987).

²⁵⁹ Stone Sweet, *supra* note 90, at 828.

²⁶⁰ See Lina Khan & David Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. __ (forthcoming 2019).

²⁶¹ See Tushnet, *supra* note 92, at 450 n.309.

Finally, weak-form review is more suitable to the nature of the online speech ecosystem. Facebook is one part of a rapidly expanding and dynamic global information environment, where norms of communication are changing on screens before our eyes and global society is still reckoning with the consequences of the hyperconnected information glut of the modern era. Weak-form review better facilitates the capacity to evolve and revise. Indeed, it seems the *only* realistic course. Entrenching decisions or making it too costly to take into account ongoing changes would cause FOB decisions to endure past their relevance. As Kadri acknowledges, “it’s true that Facebook has good reason to want some flexibility to alter its speech policies, as it constantly learns and adapts to tackle the Sisyphean task of satisfying a ‘community’ of more than 2 billion people. Binding itself to a 2018 version of its rules will surely be untenable.”²⁶² This is a fundamental point that needs to inform the entire thinking around institutions responsible with developing norms for online speech. There is still so much unknown about the new online speech environment, and further research to be done. Despite their dominance, these platforms are still a relatively young phenomenon. Platform rules should be able to be informed by new revelations. Facebook’s recent “shift” to privacy,²⁶³ for example, shows how both platforms and users can rethink their fundamental priorities and values as the ramifications of the new online speech ecosystem become apparent. Governance structures need to be able to facilitate and respond to these developments.

Of course, the FOB can and should itself revise its decisions by distinguishing or updating precedent on the basis of changed circumstances, new evidence, or developing norms. But this is an inadequate solution because the FOB cannot update decisions proactively—a *review* body is necessarily responsive to claims and disputes brought to it. One example is the updates Facebook made to its approach to violent extremism in light of the Christchurch shooting. Facebook’s policies were being updated by the day in light

²⁶² Kadri, *supra* note 233.

²⁶³ Mark Zuckerberg, *A Privacy-Focused Vision for Social Networking*, FACEBOOK (Mar. 6, 2019), <https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/> [<https://perma.cc/2W2L-L9GG>].

of its experience trying to stop the spread of a live video of the attack.²⁶⁴ It is now experimenting with audio-based technology to identify designated videos, as well as sharing URLs with other members of industry.²⁶⁵ Facebook has also revised its response to “white nationalism” and “white separatism” as part of its acknowledgment that it needs “to get better and faster at finding and removing hate from [its] platforms.”²⁶⁶ Facebook will begin removing such content, and also start directing people who post it to a nonprofit dedicated to helping people leave hate groups.²⁶⁷ Before these changes, Facebook did not prohibit this kind of content on the grounds that concepts of nationalism and separatism are “an important part of people’s identity.”²⁶⁸ These are difficult issues, and some people may argue that Facebook made the wrong call and such speech should not be censored.²⁶⁹ But the Christchurch shooting made clear the large number of people interested in spreading this kind of content, and in a way that circumvents normal platform rules.²⁷⁰ As the President of Microsoft acknowledged, “it’s clear that we need to learn from and take new action based on what happened in Christchurch.”²⁷¹ Events such as these can help inform long-standing debates about the merits of censorship versus counter-

²⁶⁴ Guy Rosen, *A Further Update on New Zealand Terrorist Attack*, FACEBOOK NEWSROOM (Mar. 20, 2019), <https://newsroom.fb.com/news/2019/03/technical-update-on-new-zealand/> [<https://perma.cc/H8PN-62DB>].

²⁶⁵ *Id.*

²⁶⁶ *Standing Against Hate*, FACEBOOK NEWSROOM (Mar. 27, 2019), <https://newsroom.fb.com/news/2019/03/standing-against-hate/> [<https://perma.cc/5PS4-ENHK>].

²⁶⁷ *Id.*

²⁶⁸ *Id.*

²⁶⁹ See, e.g., NADINE STROSSEN, *HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP* 91–94 (Inalienable Rights Series, 2018).

²⁷⁰ Liz Woolery, *Three Lessons in Content Moderation from New Zealand and Other High-Profile Tragedies*, CTR. FOR DEMOCRACY & TECH. (Mar. 27, 2019), <https://cdt.org/blog/three-lessons-in-content-moderation-from-new-zealand-and-other-high-profile-tragedies/> [<https://perma.cc/W3NZ-8CLE>].

²⁷¹ Brad Smith, *A tragedy that calls for more than words: The need for the tech sector to learn and act after events in New Zealand*, MICROSOFT ON THE ISSUES (Mar. 24, 2019), <https://blogs.microsoft.com/on-the-issues/2019/03/24/a-tragedy-that-calls-for-more-than-words-the-need-for-the-tech-sector-to-learn-and-act-after-events-in-new-zealand/> [<https://perma.cc/XXK4-C483>].

speech, which need to evolve to take account of the new empirical realities of how speech works online.²⁷²

FOB decisions should not stand in the way of such evolutions. It is easy to imagine a situation where the FOB had issued a ruling in different circumstances that would constrain Facebook's response to the Christchurch shooting: for example, a holding that sharing a URL could never be "hate speech" under the Community Standards, or that the value of "voice" prevented Facebook from censoring URL sharing. Or there may be cases where the FOB issues a decision based on an understanding of how content such as hate speech works in the offline world, which is subsequently shown to not hold in the online ecosystem. As Scanlon urges, because rights contain "a significant empirical component, our understanding of a right can always be upset by evidence that forces a change in these empirical beliefs."²⁷³ Again, such FOB decisions do not become worthless when Facebook seeks to reverse them—they will force a public dialogue about why Facebook is changing its rules or priorities.

There are many other examples of assumptions or norms that have been displaced or proven false as more is learned about speech online. Researchers have been updating their findings on whether there is a so-called "backfire effect" to labelling something as untrue.²⁷⁴ This has important ramifications for the best way to treat misinformation, an issue which platforms have been grappling with since the "fake news" scandals of 2016. Early findings about whether platforms increase "echo chambers" is also being disputed.²⁷⁵ Facebook can, and should, do much more to facilitate

²⁷² Research is being done to better understand these dynamics. Susan Benesch, *Launching Today: New Collaborative Study to Diminish Abuse on Twitter*, MEDIUM (Apr. 712, 2018), <https://medium.com/@susanbenesch/launching-today-new-collaborative-study-to-diminish-abuse-on-twitter-2b91837668cc> [<https://perma.cc/4U32-5PLW>].

²⁷³ T.M. SCANLON, *THE DIFFICULTY OF TOLERANCE: ESSAYS IN POLITICAL PHILOSOPHY* 154 (2003).

²⁷⁴ Amy Sippitt, *Does the "Backfire Effect" Exist—and Does it Matter for Factcheckers?*, FULL FACT (Mar. 20, 2019), <https://fullfact.org/blog/2019/mar/does-backfire-effect-exist/> [<https://perma.cc/9JDL-PAUB>].

²⁷⁵ Andrew Guess, *Avoiding the Echo Chamber About Echo Chambers*, MEDIUM (Feb. 12, 2018), <https://medium.com/trust-media-and-democracy/>

this kind of research by being more open with its data.²⁷⁶ But the company should not then be handicapped in its capacity to act on new findings.

It is the strong-form nature of review in the U.S. that has lead scholars such as Tim Wu to argue that First Amendment jurisprudence has been made irrelevant by the changes technology has wrought to the speech environment.²⁷⁷ Changing First Amendment jurisprudence is a heavy lift, precisely because of the difficulty of amending the U.S. Constitution or overturning Supreme Court decisions. Given how much is still unknown about online speech, and how this impacts offline lives, humility about the finality of the FOB's decisions is warranted.

3. *A Check on Users Too*

The protection of minorities, however, is one area in which the ossification of rights could actually be a significant benefit. Courts are generally an important mechanism for the protection of minority rights. This is one of the key reasons why courts are insulated from ordinary political processes, so that they are not constrained from playing this counter-majoritarian role in necessary cases. Indeed, Dworkin argued that the essence of constitutionalism is “the theory that the majority must be restrained to protect individual rights.”²⁷⁸ Such issues often arise in the context of content moderation, where many of the biggest content moderation controversies have been

avoiding-the-echo-chamber-about-echo-chambers-6e1fla1a0f39
[<https://perma.cc/Q9YG-BHR2>].

²⁷⁶ LAZER ET AL., *supra* note 185.

²⁷⁷ TIM WU, *IS THE FIRST AMENDMENT OBSOLETE?* (2017).

²⁷⁸ RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* 175 (1977). *See also* THE FEDERALIST NO. 78 (Hamilton) (“This independence of the judges is equally requisite to guard the Constitution and the rights of individuals from the effects of those ill humors, which the arts of designing men, or the influence of particular conjunctures, sometimes disseminate among the people themselves, and which, though they speedily give place to better information, and more deliberate reflection, have a tendency, in the meantime, to occasion dangerous innovations in the government, and serious oppressions of the minor party in the community.”); *West Virginia State Board of Education v. Barnette*, 319 U.S. 624, 638 (1943) (Jackson, J., writing that one’s “‘right to life, liberty and property, to free speech, free press, freedom of worship and assembly, and other fundamental rights’ may not be submitted to vote.”).

about concerns for the rights of minorities.²⁷⁹ As the UN Special Rapporteur on Freedom of Expression has noted, “[t]he vagueness of hate speech and harassment policies has triggered complaints of inconsistent policy enforcement that penalizes minorities while reinforcing the status of dominant or powerful groups.”²⁸⁰ The FOB could have an especially important role in protecting the speech of those who are otherwise likely to be drowned out on the platform or who are underrepresented in the communities of engineers and content moderators who control speech on Facebook.

Zuckerberg writes in his Blueprint that “[j]ust as our board of directors is accountable to our shareholders, [the FOB] would be focused only on our community.”²⁸¹ He is right to distinguish between *focus* on the user community and *accountability* to it. While it is important that the FOB works to further the welfare of users (as distinguished from the welfare of Facebook as a company), this is not the same thing as making decisions that simply reflect user preferences. Indeed, “[a]n independent judiciary does not take a poll before rendering its decisions.”²⁸² But there is the possibility that the FOB’s protective role is undermined through weak-form review if Facebook simply has the capacity to overrule FOB decisions that are unpopular with majorities. However, this risk does not outweigh the potential benefit of weak-form review, and the need for content moderation rules to be able to be dynamically updated. It is true that there is little to guard against Facebook overruling the FOB in the name of majoritarianism beyond Facebook’s good faith implementation of the FOB’s independent oversight and a commitment to “rule of law” style principles such that it will not overturn decisions simply on the basis that they are unpopular. But, as argued above, the very fact that content moderation decisions are

²⁷⁹ See, e.g., Schechter, *supra* note 142 (reporting on a letter from civil rights organizations, including the ACLU, criticizing Facebook: “[a]ctivists in the Movement for Black Lives have routinely reported the takedown of images discussing racism and during protests, with the justification that it violates Facebook’s Community Standards. At the same time, harassment and threats directed at activists based on their race, religion, and sexual orientation is thriving on Facebook.”); Angwin & Grassegger, *supra* note 51.

²⁸⁰ Kaye, *supra* note 14, at 10.

²⁸¹ Zuckerberg, *supra* note 1.

²⁸² Cross, *supra* note 135, at 559.

controversial and will tend to upset significant portions of users is part of the reason Facebook is setting the FOB up. This desire to outsource controversy suggests that mere unpopularity of a decision would not be a reason for Facebook to intervene because then it would be undermining its own project to distance itself from the controversies. In this sense, the lack of democratic accountability or elections of Facebook's management could provide further protection from simple majoritarianism.

B. Public Reasoning

Public reason-giving is the defining characteristic of the FOB. Facebook already has internal teams that continually review content moderation standards and consult outside experts.²⁸³ The utility of the FOB is not merely in having experts review Facebook's decisions and coming to a "right" outcome. If it were, the internal processes would be sufficient. Instead, the FOB is intended to serve a function separate and additional to having experts weigh in on Facebook's rules. The core value offering of the FOB is the issuing of public explanations for rulings.²⁸⁴ As Noah Feldman wrote in his original white paper proposing the FOB:

The advantage enjoyed by real-life constitutional courts is that they openly address difficult cases, and so derive credit and legitimacy from being principled. They make mistakes, and correct them. Their rules evolve with changing technology and ideas. And instead of blaming them for this, we mostly validate their efforts. . . . Right now, the platforms are already doing plenty of balancing work. But they aren't doing it transparently or in a legal-logical way. Changes are greeted with outrage rather than respectful engagement. All that could change if the

²⁸³ *Facts About Content Review on Facebook*, FACEBOOK NEWSROOM (Dec. 28, 2018), <https://newsroom.fb.com/news/2018/12/content-review-facts/> [https://perma.cc/UM2P-JUHE] (Facebook noting it holds "a global forum held every two weeks where we discuss potential changes to our policies. It includes experts from around the world with deep knowledge of relevant laws, online safety, counter-terrorism, operations, public policy, communications, product, and diversity.").

²⁸⁴ Casey Newton, *Facebook Will Create an Independent Oversight Group to Review Content Moderation Appeals*, VERGE (Nov. 15, 2018), <https://www.theverge.com/2018/11/15/18097219/facebook-independent-oversight-supreme-court-content-moderation> [https://perma.cc/P4QJ-96QT]. See also DRAFT CHARTER, *supra* note 2, at 5.

platforms provided a forum for argument, openly considered opposing views, and announced the reasoning behind their decisions on a case-by-case basis.²⁸⁵

1. *The Purpose of Public Reasons*

What exactly, then, is the purpose of public reasoning? An obvious answer might be that in an individual case adequate reasons are an incident of due process and fairness, by allowing a user to know the reasons for a decision (especially an adverse one) involving their freedom of expression on the platform.²⁸⁶ This is why many common law jurisdictions have a judicial duty to give reasons as an incident of due process²⁸⁷ and all constitutional courts of Europe are obliged to give reasons for their decisions.²⁸⁸ This is also why the Santa Clara Principles, a set of minimum requirements for transparency and accountability for content moderation proposed by a group of leading civil society organizations, academics, and experts, requires that users be given “a statement of reasoning sufficient to allow the user to understand the decision.”²⁸⁹ But, as has been noted repeatedly throughout this article, the due process offerings of the FOB will be limited due to the small fraction of cases it will adjudicate. Furthermore, affected users could be given reasons without them being made public, which would also diminish privacy concerns in releasing public opinions. Therefore, there must be other, more general, rationales for public reason-giving.

As a practical matter, giving reasons assists in creating a consistent body of case law by allowing the FOB and users to understand the basis on which a particular decision is made and distinguish or apply it in future cases. Stone Sweet describes giving

²⁸⁵ Harris, *supra* note 138, at 138.

²⁸⁶ See, e.g., Court of Appeal in *Flannery v Halifax Estate Agencies Ltd* [2000] 1 WLR 377, 381 (“The duty is a function of due process, and therefore of justice. . . . [F]airness surely requires that the parties especially the losing party should be left in no doubt why they have won or lost. . . . [A] requirement to give reasons concentrates the mind; if it is fulfilled, the resulting decision is much more likely to be soundly based on the evidence than if it is not.”).

²⁸⁷ H.L. Ho, *The Judicial Duty to Give Reasons*, 20 LEGAL STUD. 42 (2000).

²⁸⁸ STONE SWEET, *supra* note 8, at 56 n.15.

²⁸⁹ *The Santa Clara Principles on Transparency and Accountability in Content Moderation*, <https://santaclaraprinciples.org> [<https://perma.cc/NG9H-KX4Q>] (last visited Sept. 23, 2019).

defensible reasons that are accepted as a precedential interpretation of constitutional meaning as “necessary conditions for the emergence of effective review systems.”²⁹⁰ But, reasons do not need to be available to the public-at-large for moderators to apply them. So, the goal of reason-giving is still more general than that.

Fundamentally, the FOB is a response to calls for decisional transparency and a “global dialogue” with users and stakeholders about the impacts and justifications of Facebook’s rules.²⁹¹ Platforms generally have been increasingly criticized as having opaque decision-making that interferes with their obligations of clarity, specificity, and predictability.²⁹² Public reasoning, then, is about facilitating this dialogue and legitimizing Facebook’s exercise of power over its users. As Rawls argued, in a pluralistic society where there will always be disagreement about what rule is best, the exercise of power over those who disagree with decisions is *only* legitimized through public reasoning that proceeds in a way people might be expected to respect.²⁹³ Fallon notes that when decision-makers rely on reasons that reasonable people would acknowledge as fair (and not, for example, idiosyncratic or partisan), this creates legitimacy even if people might reach different ultimate judgments.²⁹⁴ Therefore, the goal is emphatically not to create collective agreement, but to allow for reasoned disagreement. Tom Tyler’s research has shown that people’s judgments of legitimacy do not depend primarily on their obtaining favorable outcomes, but are more strongly influenced by the processes and procedures authorities use, including whether they afford participation, demonstrate impartiality, and show respect for people’s interests as worthy of consideration.²⁹⁵ Therefore, public reasoning is crucial to the FOB’s central goal of creating legitimacy for Facebook’s rules.

²⁹⁰ Stone Sweet, *supra* note 90, at 825.

²⁹¹ ASH ET AL., *supra* note 16, at 13; Kaye, *supra* note 14.

²⁹² Kaye, *supra* note 14.

²⁹³ RAWLS, *supra* note 28, at 217; *see also* FALLON, *supra* note 76, at 12; Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 230 (2004).

²⁹⁴ FALLON, *supra* note 76, at 128.

²⁹⁵ Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 3 CRIME AND JUSTICE 283 (2003); Facebook Data Transparency Advisory Group, *supra* note 88, at 34.

Until now, Facebook has exercised substantial power over not only individual speech, but entire societies' public discourse, while giving little justification for the way in which it does so. The FOB is a step towards changing this. Just as constitutional courts can provide "a focal point for a new rhetoric of state legitimacy" during democratic transitions,²⁹⁶ so too might the FOB provide a forum for working through the principles in this new era of online governance.

Two opposing examples are illustrative of the benefits of public reasoning. In the first example, a lack of public reasoning delegitimized a decision even for those that agreed with it. In the second, reasoning helped the affected user come to accept a decision they still thought was wrong.

The first example is the controversy over Alex Jones' presence on Facebook. Leaked emails show the difficulty internal executives were having deciding whether a particular post violated Facebook's Community Standards.²⁹⁷ In one email, an executive referred to the fact that the number of comments on the post that violated the Community Standards "did not meet the threshold for deletion."²⁹⁸ UK executives then pointed out local context, noting that the image is famous in the UK and "widely acknowledged to be anti-Semitic" there, justifying deletion.²⁹⁹ These are the kinds of arguments the FOB could help ventilate. The "threshold" referred to is not defined in any public document. Whether or not this is a legitimate way of determining whether content should be removed, the fact that it is applied in a non-transparent way by executives that seem to disagree about whether the standard is met deprives it of legitimacy as a rule that users can reasonably agree to be bound by. The FOB might ultimately decide to apply the same standard—again, the benefits of the FOB are not necessarily to reach different or "better" outcomes. But if the FOB did decide to apply this test, it would need to explain

²⁹⁶ Stone Sweet, *supra* note 90, at 827.

²⁹⁷ Jake Kanter, *Leaked emails reveal Facebook's intense internal discussion over Alex Jones' "anti-Semitic" post on Instagram*, BUSINESS INSIDER (Mar. 28, 2019), <https://www.businessinsider.com/facebook-emails-reveal-discussion-about-alex-jones-instagram-account-2019-3> [<https://perma.cc/H3FR-EFV9>].

²⁹⁸ *Id.* Infringing comments amounted to about 4% of the total number of comments on the post.

²⁹⁹ *Id.*

this decision publicly which would allow for greater consistency in its application as well as contestation about the appropriateness of the threshold in the first place. Similarly, public reasoning can excavate local context to inform users unfamiliar with it why a post has special meaning in different circumstances. But the importance of public reasoning is most evident in the reaction to the ultimate decision to take down Jones' pages altogether. Although many commentators had been calling for his removal for some time, the response was mixed. Few defended Jones, but there was frustration with the way Facebook executed the ban and the lack of transparency around the reason or timing.³⁰⁰ Without public reasoning, even those that agreed with the decision thought it was illegitimate.

The second example is from a different platform but illustrates the benefits of giving understandable rationales. David Neiwert's Twitter account was suspended when he changed his profile picture to the cover of his book about the alt-right, which included KKK

³⁰⁰ April Glaser, *Why Facebook's Latest Ban of Alex Jones and Company Was So Underwhelming*, SLATE (Apr. 24, 2018), <https://slate.com/technology/2019/05/facebook-alex-jones-ban-underwhelming.html> [https://perma.cc/S5HU-5A2U] ("Facebook has the power to punish wrongdoers, as it did on Thursday. But we don't know its full rationale for doing so, nor do we know who will be next."); Bret Stephens, *Facebook's Unintended Consequence*, N.Y. TIMES (May 3, 2019), <https://www.nytimes.com/2019/05/03/opinion/facebook-free-speech.html> [https://perma.cc/PY7M-DFGM] ("The deeper problem is the overwhelming concentration of technical, financial and moral power in the hands of people who lack the training, experience, wisdom, trustworthiness, humility and incentives to exercise that power responsibly. . . . the decision to absolutely ban certain individuals will always be a human one. It will inevitably be subjective."); David French, *Why Facebook's Bans Warrant Concern*, NAT'L REV. (May 3, 2019), <https://www.nationalreview.com/corner/why-facebooks-bans-warrant-concern/> [https://perma.cc/XX9Q-V92E] ("This means a person can potentially face social-media bans even if they comply with every syllable of the company's speech rules on the company's platform. The potential for abuse is obvious, as is the potential chilling effect"); Emily Stewart, *Facebook bans Alex Jones, Infowars, Louis Farrakhan, and others it deems "dangerous,"* VOX (May 2, 2019), <https://www.vox.com/recode/2019/5/2/18527357/facebook-bans-alex-jones-louis-farrakhan-infowars> [https://perma.cc/WUX9-KSA8] ("It's not clear why Facebook is doing this now, but pressure for it to take action has been mounting for quite some time, and the decision is probably at least in part an effort to get some positive PR.").

hoods.³⁰¹ Neiwert thought the suspension was wrong and refused to change the picture. From his perspective, the image was about analyzing hate, not promoting it. Representatives from Twitter reached out and explained that although they appreciated the distinction he was making, the company takes a no-tolerance stance on such images in profile pictures because they are more prominently displayed on the site. Neiwert wrote that the conversation “was cooperative and [they were] genuinely interested in my input. These Twitter officials were able to persuade me, at least, that they very much share my concerns.”³⁰² Neiwert still disagreed with the decision, but once he understood the reasoning he agreed to change his profile picture and his account was reinstated.

These examples illustrate the ultimate hope for the FOB. As Feldman summarizes:

Some controversy and disagreement over the [FOB’s] decisions is inevitable. But even when it occurs, it will come in the context of the public understanding that Facebook is publicly and responsibly grappling with balancing values in cases that have no simple right answers. This repeated engagement should produce legitimacy for the decision process, and a new narrative for Facebook’s engagement with these problems.³⁰³

Another benefit of public reasoning is its ability to help in norm-setting. The FOB can provide contestation and explanation of norms in a more public forum. This in turn might make more community members aware of the rules which helps generate compliance. Grimmelman describes the importance of this aspect of moderation, saying that moderation’s “most important mission is to create strong shared norms among participants. . . . Moderators can influence norms directly by articulating them. They can do this either in general, with codes of conduct and other broad statements of rules, or in specific cases by praising good behavior and criticizing bad.”³⁰⁴

³⁰¹ David Neiwert, *Why I Ended My Fight With Twitter—For Now*, DAILY KOS (June 25, 2019), <https://www.dailykos.com/story/2019/6/25/1867213/-Why-I-ended-my-fight-with-Twitter-for-now> [<https://perma.cc/4YTW-3R8R>].

³⁰² *Id.*

³⁰³ Harris, *supra* note 138, at 145.

³⁰⁴ James Grimmelman, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 61–2 (2015).

Matias similarly found that the visibility of the rules of online communities substantially increases compliance and overall participation in the community.³⁰⁵ By allowing greater visibility and participation in content moderation decisions through public reasoning, the FOB can embed the process of rule formation in a broader community and help norms be formed and tested.

Public reasoning is also an important constraint on the FOB itself. Although it should be independent, this does not mean the FOB should be unconstrained. Giving principled reasons is the primary way that judges can be held accountable for *their* exercise of power.³⁰⁶ As Mark Tushnet writes, “the desideratum is not judicial independence alone but rather judicial independence coupled with accountability to law.”³⁰⁷ He explains that “without the latter, independent judges can act arbitrarily and so anticonstitutionally.”³⁰⁸ The FOB will need to develop a form of discourse, centered around norms, that distinguishes it from a mere political institution.³⁰⁹ It is the norms and constraints of principled reason-giving that will reassure users that they are not merely “exchanging one set of tyrants for another.”³¹⁰ This legitimacy-creating constraint is especially important when ambiguity is common. Accountability to law does not mean demonstration that the law *requires* a particular result. Indeed, as Rawls says, “public reason often allows more than one reasonable answer to any particular question.”³¹¹ What is important is not convincing people that the reasoning methods employed offer determinate outcomes but, as Roux has observed, that decision-makers should use

³⁰⁵ J. NATHAN MATIAS, PROCS. OF THE NAT’L ACAD. OF SCIENTISTS, PREVENTING HARASSMENT AND INCREASING GROUP PARTICIPATION THROUGH SOCIAL NORMS IN 2,190 ONLINE SCIENCE DISCUSSIONS (2019).

³⁰⁶ Roderick A. Macdonald & Hoi Kong, *Judicial Independence as a Constitutional Virtue*, in THE OXFORD HANDBOOK OF COMPARATIVE CONSTITUTIONAL LAW 831, 839 (Michel Rosenfeld & András Sajó eds., 2012).

³⁰⁷ Tushnet, *supra* note 92, at 418–19. *See also* GUARNIERI & PEDERZOLI, *supra* note 189, at 45; Cross, *supra* note 135, at 558.

³⁰⁸ Tushnet, *supra* note 92, at 419 n.145; *see also* GUARNIERI & PEDERZOLI, *supra* note 189, at 196.

³⁰⁹ GUARNIERI & PEDERZOLI, *supra* note 189, at 196.

³¹⁰ RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 6 (1993).

³¹¹ RAWLS, *supra* note 28, at 240.

“methods that have come to be seen by the legal community and the broader legal culture as legitimate.”³¹² Public reasoning is the way in which the FOB can show its work in this regard.

2. *The Challenge of Giving Acceptable Reasons*

There are four particular aspects of the FOB’s context that will make its use of public reasoning to create legitimacy uniquely challenging. First, neither the FOB itself nor the community it speaks to are socialized in a particular style of discourse, such as legal reasoning. The FOB’s reasons and those that read them will not draw on what Llewellyn called law’s “steadying factors” which provides accepted doctrinal techniques for legal argument.³¹³ Judicial legitimacy is itself at least partially rooted in adherence to this craft.³¹⁴ Of course, some contest the constraining power of these norms and even their existence at all.³¹⁵ But most accept that legal decisions are something different to mere politics.³¹⁶ So while there is some degree of choice, most constitutional scholars for example agree there is still something distinctively legal inherent in the process of constitutional construction.³¹⁷ Setting aside the fundamental debate over legal realism, it is sufficient for present purposes to note that finding a mode of discourse that can be accepted as legitimizing in Facebook’s vast and diverse community will only be harder when not tied to a previously existing set of professional norms.

This is exacerbated because, second, there are no accepted global norms to guide the substantive issues. Discussing the diversity of views on freedom of expression, Facebook’s Head of Global Policy Management has written:

³¹² Theunis Roux, *Reinterpreting the Mason Court Revolution: A Historical Institutional Account of Judge-Driven Constitutional Transformation in Australia*, 43 FED. L. REV. 1, 24 (2015).

³¹³ KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 5, 19–23 (1960). Ginsburg notes that professional norms of fidelity to law are a kind of internalized ideology that help reduce the agency costs of using judges as agents to monitor administrative structures. *See* Ginsburg, *supra* note 119, at 65.

³¹⁴ Fallon, *supra* note 205, at 1826.

³¹⁵ *Id.* at 1826–27.

³¹⁶ FALLON, *supra* note 76, at 2.

³¹⁷ Landau & Dixon, *supra* note 81, at 23.

Laws differ from country to country, not only in their text, but also in how their enforcement is pursued by authorities and how they are applied by courts. Even within the confines of one country, and even within common law jurisdictions, courts differ in their interpretation of laws. To complicate matters, laws evolve. New laws are passed, old laws are amended, and courts invalidate some laws altogether. . . . [E]ven assuming one government's laws represent the social norms of its people, those laws will not represent the norms of every person on a global social media service.³¹⁸

Adopting international law standards does not simplify the matter. These are not universally endorsed.³¹⁹ Nor are they themselves stable—for example, “international doctrine and practice relating to prohibition of hate speech remain uneven.”³²⁰ Procedural elements of resolving disagreement are not so easily divorced from the underlying substance.³²¹ Finding arguments that can reasonably be regarded as acceptable across such diverse substantive traditions will be especially challenging.

Third, compounding the difficulties, Facebook does not have an underlying rationale for free speech on its platform to justify its choices. As Adrienne Stone notes, even in nation states, “identifying the value or set of values underlying any single constitutional system of freedom of expression is likely to be difficult.”³²² For Facebook, it is even more difficult. What is the purpose of speech on Facebook? The simple answer that it generates revenue by capturing attention is not one that will help satisfactorily justify outcomes in difficult cases. But the three most common rationales from other traditions do not fit easily in the context of Facebook: facilitating self-government, the search for truth, or respect for individual autonomy. This idea that freedom of expression is valuable because of its capacity to promote democratic self-government is the most widely

³¹⁸ Bickert, *supra* note 31, at 260–61.

³¹⁹ See, e.g., Amal Clooney & Philippa Webb, *The Right to Insult in International Law*, 48 COLUM. HUM. RTS. L. REV. 1 (2017) (critiquing permissible restrictions on freedom of expression under international law).

³²⁰ Cleveland, *supra* note 59, at 225.

³²¹ JEREMY WALDRON, *LAW AND DISAGREEMENT* 157 (1999).

³²² Adrienne Stone, *The Comparative Constitutional Law of Freedom of Expression*, in *COMPARATIVE CONSTITUTIONAL LAW* 406, 416 (Tom Ginsburg & Rosalind Dixon eds., 2011).

adopted in modern legal systems,³²³ but seems inappropriate in the context of Facebook. Facebook is not a democracy—it is a business.³²⁴ It does not rely on popular will in setting its rules (a brief experiment with a limited such system in 2009 failed due to poor voting levels).³²⁵ It has expressly disavowed any conception of its platform as facilitating the search for truth.³²⁶ In any event, its algorithm manipulates the “marketplace of ideas” which is metaphorically said to lead to truth’s revelation. This leaves the third dominant justification for freedom of speech: that based on individual autonomy.³²⁷ Perhaps this is the most fitting understanding of Facebook’s justification for free speech, but still sits uneasily with the fact that Facebook regularly censors speech. It is therefore difficult to pin down any underlying theory behind Facebook’s content moderation system which could inform the FOB’s work, and which could satisfy those that disagree with the FOB’s decisions, that they are based on some more fundamental principle. As discussed above, Facebook’s statement of values might mitigate this, but only to a certain extent. A decision to prioritize “safety” would give a clear guiding principle in many individual decisions. But if Facebook decides its central value is “voice,” this will not necessarily make trade-offs easier. Whose

³²³ *Id.* at 414.

³²⁴ Pozen, *supra* note 32 (explaining that there is a “tension between Facebook’s seemingly sincere concern for free speech values and its explicit aspiration to make users feel socially safe and ‘connected’ [and hence to maximize the time they spend on the site], a tension that is shaped by market forces but ultimately resolved by benevolent leader and controlling shareholder Zuckerberg.”).

³²⁵ Adi Robertson, *Facebook used to be a democracy — but nobody voted*, VERGE (Apr. 5, 2018), <https://www.theverge.com/2018/4/5/17176834/mark-zuckerberg-facebook-democracy-governance-vote-failure> [<https://perma.cc/F4TN-AUQM>].

³²⁶ Andrew Liptak, *Mark Zuckerberg warns about Facebook “becoming arbiters of truth,”* VERGE (Nov. 13, 2016), <https://www.theverge.com/2016/11/13/13613566/mark-zuckerberg-facebook-misinformation-hoax-media> [<https://perma.cc/W8VR-QFHY>]; Arjun Kharpal, *Sandberg: Facebook Doesn’t Want to Be the ‘Arbiter of the Truth,’* DAILY BEAST (Apr. 24, 2017), <https://www.thedailybeast.com/cheats/2017/04/24/facebook-not-the-arbiter-of-the-truth> [<https://perma.cc/EL85-A369>].

³²⁷ Stone, *supra* note 322, at 413–14; see FREDERICK F. SCHAUER, *FREE SPEECH: A PHILOSOPHICAL ENQUIRY* (1982).

voice should be given priority, for example? Justifying decisions to prioritize voice without a thesis of the underlying purpose of voice will limit the extent to which justifications can legitimize choices.

These are the challenges for the FOB fulfilling its role as an “exemplar of public reason.”³²⁸ Nevertheless, it is a central and important goal for the institution. As more expression occurs in these online spaces, it is unavoidable that there will need to be some body that performs this role of rationalizing contested and contestable decisions about how to regulate important public discourse. For now, the FOB is the first attempt at such an institution. It might be a canary in the coal mine of the future of online governance. No doubt, much will be learned from its successes and, perhaps inevitably, blunders.

VI. CONCLUSION

The FOB represents an important innovation, and a welcome attempt to disperse the enormous power over online discourse held by Facebook. More importantly, however, it will serve to make that power more transparent and legitimate by facilitating dialogues around how and why Facebook’s power is exercised in the first place. This is a more modest goal than becoming an independent source of universally accepted free speech norms, but it is still incredibly ambitious for an institution that is breaking new ground. As Stone Sweet writes about constitutional courts in regime transitions:

The ultimate measure of legitimacy for any [constitutional court] may well be its success at helping the polity construct a new constitutional identity—a massive undertaking. . . . [A]s Scheppele writes, a [constitutional court] is often ‘the primary mechanism’ for organizing the transition away from the former ‘regime of horror’ to constitutional democracy. Insofar as [constitutional courts] are successful, the legitimacy of the constitution, as a basic framework for the exercise of public authority, will become indistinguishable from the regime’s political legitimacy.³²⁹

For the best chances of success, Facebook and the FOB should be humble and acknowledge the very real limitations that such a

³²⁸ RAWLS, *supra* note 28, at 231.

³²⁹ Stone Sweet, *supra* note 90, at 829.

body will face. As a Facebook representative notes, “We are very much at the beginning of this process—it has not been done before”³³⁰ There is no true model to base the new body on. Deciding what the FOB will be includes accepting what it cannot be—this includes acknowledging that it cannot be a way to bring due process to any but the smallest fraction of content moderation decisions. Nor can it authoritatively resolve clashing ideas of freedom of expression. But, if done right, the FOB may be able to bring a greater sense of legitimacy and acceptance to Facebook’s content moderation ecosystem. This is a massive undertaking, and will require a healthy dose of “constitutional luck.”³³¹ History teaches us that ultimately many variables for constitutional success are beyond the ability of designers to control, and success or failure is necessarily contingent. This may be unsatisfying. But the question should not be whether the FOB is inevitably a perfect institution or an ideal-type of due process and transparency that will bind Facebook to stringent human rights standards. The question for now is whether it is better than the alternative: the current haphazard, opaque process that draws inspiration from Kafka more than Kelsen. By that standard, the FOB shows real promise.

³³⁰ Archibong, *supra* note 117.

³³¹ Justin Blount et al., *Does the Process of Constitution-Making Matter?*, in COMPARATIVE CONSTITUTIONAL DESIGN 43 (Tom Ginsburg ed., 2012) (citing Kim Lane Scheppele, *A Constitution between Past and Future*, 49 WM. & MARY L. REV. 1377, 1387 (2007)).

