

Recommendations for the Facebook Content Review Board

2018-19 PRACTICUM RESEARCH TEAM:

Shaimaa BAKR, Ph.D. Electrical Engineering '20 Madeline MAGNUSON, J.D. '20
Fernando BERDION-DEL VALLE, J.D. '20 Shawn MUSGRAVE, J.D. '21
Isabella GARCIA-CAMARGO, B.S.'20 Ashwin RAMASWAMI, B.S. '21
Julia GREENBERG, J.D. '19 Nora TAN, B.S. '19
Tara IYER, B.S.'19 Marlena WISNIAK, LL.M. '19
Alejandra LYNBERG, J.D. '19 Monica ZWOLINSKI, J.D. '19

INSTRUCTORS:

Paul BREST, Faculty Director, Law and Policy Lab
Daniel HO, William Benjamin Scott and Luna M. Scott Professor of Law
Nathaniel PERSILY, James B. McClatchy Professor of Law
Rob REICH, Faculty Director, Center for Ethics in Society

TEACHING ASSISTANT:

Liza STARR, J.D./M.B.A. '21

POLICY CLIENT:

Project on Democracy and the Internet, Stanford Center on Philanthropy and Civil Society

**SPRING
2019**



Contents

Executive Summary	3
Introduction	8
I. The Need for Governing Principles.....	9
II. Structure and Membership.....	10
A. Selection Process for Board Members.....	10
B. Membership Criteria.....	12
C. Structure of the Board.....	15
III. Case Selection	18
A. Cases Eligible for Board Review	19
B. Petitioners	20
C. Selection Process.....	23
D. Information Required about Facebook’s Appeals Process.....	29
IV. Procedures.....	31
A. Overview: Timeline	31
B. Initial Disclosures.....	32
C. Deliberation.....	33
D. Decision	36
V. Implementation of Board’s Decisions	39
A. Decision Scope	39
B. Remedies	40
C. Decision Authority on Facebook.....	41
D. Precedential Value of Board Decisions.....	43
E. Monitoring	44
F. Annual Report.....	45
Conclusion.....	46
Appendix.....	47
A. Example Hate Speech Petition	47

Executive Summary

In order to provide greater transparency and accountability for enforcement of its Community Standards, Facebook has called for the establishment of an independent board to oversee content decisions.¹ Following a period of consultation with NGOs and other interested parties around the world, Facebook has asked for public comment and input on the concept and structure of the board. In partnership with its policy client, the [Stanford Project on Democracy and the Internet](#), the [Stanford Law and Policy Lab](#) engaged in research to inform the design of a social media oversight board, within the broad outlines of the [draft charter](#) that Facebook issued for public comment.

The recommendations presented here reflect the work of a team of Stanford law and engineering students, who, over the spring 2019 term, engaged in qualitative interviews with Facebook personnel, representatives from NGOs, academics, and legal experts to assess tradeoffs and prioritize options across different adjudicatory and regulatory models. The report assesses different design features for an oversight board for Facebook. It discusses the need for governing principles for any oversight board, the structure and membership of such a board, its case selection process, procedures for the board to review cases and issue decisions, and the board's authority to implement its decisions.

The Need for Governing Principles

If the board is to have any independence and authority, its responsibility cannot be merely to interpret Facebook's internal Community Standards. The reason is that the Community Standards are nearly continuously revised by Facebook's Content Policy team. Any interpretation of a provision in the Community Standards is, therefore, subject to Facebook reversal, rendering all board decisions potentially advisory. While we take no position on the appropriate source of authority (whether a "Constitution" for Facebook or international human rights law), we believe serious thought should be given to an external source of governing principles that the board applies in its decisions.

Although we do not make specific recommendations on the source of authority, we agree with the general principles as spelled out by Facebook. First, the board should be independent. Independence requires that no other entities, including Facebook itself, exert inappropriate or unwarranted interference in the board's operations, composition, and decision-making. Second, the board should be publicly accountable and transparent in its decision-making process. The board should ensure transparency in its decisions while providing requisite privacy protections. Third, the board should reflect the wider society it serves. Facebook recognizes this point, remarking that "the 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."²

1 See Facebook, *Draft Charter: An Oversight Board for Content Decisions*, January 28, 2019, <https://perma.cc/6T65-LWKB>. The research from this Stanford Law and Policy Lab report is issued independently of Facebook's June 27, 2019, report, *Global Feedback & Input on the Facebook Oversight Board for Content Decisions*, <https://perma.cc/RA6R-9TG8>.

2 Ime Archibong, "Strengthening Safety and Encouraging Free Expression: Facebook's Goals for a New Content Oversight Board," Facebook Newsroom (blog), March 26, 2019, <https://perma.cc/FXB6-W8X5>.

Structure and Membership

To be effective, the board must be global, diverse, and sufficiently well-staffed to select and consider a diverse array of cases.

- **Membership.** Facebook has stated its intention to establish a forty-member, part-time board to handle content review decisions. We believe a part-time board of that size will not be sufficient to handle case selection, thorough consideration of the evidence, deliberation, and decision writing at any meaningful scale. We hence recommend that Facebook consider making board positions full-time, with staggered terms, and/or increasing the number of part-time board members.
- **Inclusive Representation.** Diversity in board membership will ensure that a wide range of perspectives — including national origin, ethnicity, or religion — is taken into account in the board’s decisions. To ensure that the board reflects the diversity of Facebook users, the board membership should reflect, in some proportionality, Facebook’s user population, while not allowing any one set of voices to dominate the rest.
- **Appointment Process.** To ensure board independence, Facebook may want to refrain from single-handedly appointing board members and instead constitute an independent selection committee. This committee should also focus on ensuring diversity according to criteria of geography, expertise, race, religion, and gender.
- **Staff.** Given the complexity and caseload, the board will need a substantial permanent staff to aid with case selection, case preparation, research, writing, and monitoring.
- **Funding.** The board cannot be institutionally independent if Facebook can cease funding it at any point. Facebook should explore mechanisms to guarantee funding for the board, such as an independent trust fund or endowment.

Case Selection

To maintain credible independence from Facebook, the board must be able to select its own cases. This process should include:

- **Petitioning for Review.** Any interested party who has exhausted his or her ability to file appeals internal to Facebook and wants to appeal a content decision to the board should be able to petition the board for review.³ This includes posters of content, flaggers of content, targets of content, and interested third parties such as Facebook itself, non-governmental organizations (NGOs), other companies, and governments. The petition forms should be structured, requiring petitioners to expend some effort and also eliciting relevant contextual information to guide the board through selection and decision.

³ Our understanding of the current internal appeals system is that exposure to the content is required to be able to petition for review. The exhaustion requirement may need to be relaxed or the internal petition process expanded in instances where parties whose interests are clearly implicated (e.g., an NGO that is not a Facebook user or a user who is depicted in but not exposed to content) are unable to exhaust internal mechanisms.

- **Case Selection.** A three-tiered system can help cope with the vast scale of disputed content decisions. “Elevation Teams,” organized by region and content category and comprised of full-time staff members, can sort through the large pool of petitions to present a subset to a “Selection Panel” twice each month for consideration. The Selection Panel, composed of board members, can consider the cases nominated by the Elevation Teams, alongside a random (or stratified random) sample of cases, and decide which cases to select for review by a “Decision Panel.”
- **Selection Criteria.** To select cases, the Elevation Teams and Selection Panels should follow established criteria, prioritizing cases that: (1) raise important undecided issues, (2) reveal considerable variance between moderators, (3) involve potential miscarriages of justice, and (4) conflict with board precedent and/or governing principles. Selectors should also ensure that cases reflect the range of issues along geographic region, types of petitioners, and types of violations.
- **Information Sharing.** Given the vast quantities of content removed by Facebook and appealed by users, Facebook will need to give the board access to existing internal decision-making processes for appeals. In order to apply the selection criteria listed above, board selectors need to be able to access and sort petitions based on metrics such as how often particular pieces of content are shared, flagged, and removed; selectors will also need to understand how Facebook content moderators make their decisions and utilize do-not-post rules and other internal Facebook guidelines.
- **Appeals Eligibility.** Facebook should ensure that adequate appeals mechanisms are in place for all types of content and that the board has authority to review those decisions. While some types of content (e.g., copyright claims) may be generally ill-suited to board review, that content should still be within the board’s purview, given that, for example, pretextual copyright claims could be used to shield an otherwise improper takedown from further review.

Procedures

Clear and effective procedures are essential for the legitimacy of an oversight board, including:

- **Case Review.** Cases should be heard by three-member panels where at least one panel member represents the region in question. In cases of exceptional importance, en banc review by a panel of nine (randomly selected) board members may be granted after a majority vote by the en banc panel.
- **Input and Evidence for Cases.** In addition to statements from direct parties, cases may include expert briefs on relevant geopolitical, social, and cultural contexts, national law, and international human rights law.
- **Opinions.** Opinions should describe the panel’s reasoning and identify criteria for the decision. Any panel member may file a dissenting opinion. Opinions should always be issued in a timely fashion, but the board may also establish mechanisms for expedited review of urgent matters.
- **Publication.** Pending cases, membership, and board procedures should be posted on the board website. Opinions should be posted, where appropriate, but may be redacted for privacy and other reasons. To protect board member deliberation, specific panelists assigned to individual cases should not be identified.

Implementation of Board's Decisions

To be effective, the board should be able to deploy a wide range of remedies and monitor content moderators and algorithms for compliance with its decisions.

- **Remedial Authority.** The board should have the authority to review disputed content for content moderation errors and, as necessary, issue decisions that directly address Facebook content moderation policies, practices, and Community Standards.
- **Precedent.** The board should have authority to determine which of its content-related decisions have precedential authority and specify the scope of its precedents, enhancing the board's flexibility to cabin decisions that are context-specific.
- **Orders to Revise Community Standards, Training, and Algorithms.** Given that the underlying problems of content moderation may emanate from the Community Standards, training, or algorithmic decisions, the board's power should extend beyond isolated content decisions. Where appropriate, the board should have the power to order Facebook to revise the Community Standards, modify training, and update algorithmic content decisions. Unlike orders with respect to specific content, these board orders should be rebuttable: Facebook may choose not to implement the board's order, but it must provide a public, written rationale of refusals to comply with an order.
- **Monitoring.** Effective and legitimate oversight can be achieved only if Facebook adheres to the board's decisions. The board should monitor Facebook's implementation of its decisions in three areas: (1) updates to the Community Standards, (2) content moderator adherence to board decisions, and (3) algorithmic adherence to board decisions.
- **Annual Report.** Board members will develop a unique perspective on Facebook's rapidly evolving content moderation system. Based on its experience and perspective, the board may issue an annual report making recommendations for reform.

Conclusion

We applaud Facebook for its creative proposal to address the problems of content moderation. As other social media platforms contemplate reforming their content moderation policies, Facebook has the unique opportunity to lead by example. If designed poorly, the board runs the risk of serving as a quasi-judicial rubber stamp on a system plagued by deep structural problems. If implemented well, the board may help to institute genuine reform and systematic improvement, based on transparency and accountability.

Introduction

In the space of twenty years, the infrastructure of free expression and association has undergone a profound transformation. The digital revolution has given billions of people the opportunity to express themselves in words, images, and videos to billions of others. Yet this revolution has come at a price: Social media platforms have unintentionally become gatekeepers of free expression worldwide. With social media companies increasingly under fire for their regulation of speech online, new solutions for content moderation are required.

Facebook is now incubating one potential solution: the creation of a social media oversight board. On November 15, 2018, Mark Zuckerberg announced that Facebook would create an independent governance and oversight body to oversee content moderation decisions.⁴ This independent oversight board aims to enhance the company's existing means of moderating content and alleviate the burden on internal content officers to manage complex instances of arguably impermissible speech.

As part of the process of developing this board, Facebook opened a call for public comment and input. That charge led to a Draft Charter and a report on Global Feedback & Input on the Facebook Oversight Board for Content Decisions.⁵ This report presents research by the Stanford Law and Policy Lab on how to design such a board. This research reflects the work of a team of Stanford law and engineering students, who, over the spring 2019 term, studied the problem and asked: If one were designing an oversight board, what should it look like?

This independent research contributes broadly to work underway with the Stanford Project on Democracy and the Internet, which served as this project's policy client. While the recommendations are specific to Facebook, the research seeks to inform concerns across social media platforms as they seek methods to manage content decisions. The research team engaged with Facebook officials and experts across the field to learn about the problem, but did not have full access to Facebook's internal data or the considerations facing its existing cohorts of content moderators, product designers, or algorithm developers. The team of student researchers developed consensus positions, which are presented in this report, but those views do not reflect the individual opinions of each student and advisor.

Thus, by necessity, our proposed design is the result of difficult choices. If the board is nimble and responsive to crises, its deliberative ability may suffer. If the board seeks to maintain consistent standards for Facebook's worldwide user base, local context will invariably get lost. And the more control the board has over Facebook's content moderation decisions, the greater the risk of open conflict between the board and Facebook. Facebook must weigh these tradeoffs, among others, and will need to make compromises. Although the board cannot satisfy all of these competing and contradictory demands, its success will depend on its perceived legitimacy and the efficiency with which it handles a diverse set of cases arising from content decisions made throughout the world.

4 Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement," Facebook, November 15, 2018, <https://perma.cc/4WBK-Z33P>.

5 Facebook, *Draft Charter*; Facebook, *Global Feedback*.

This report is organized into five sections. Section I briefly describes the need for governing principles for the board, beyond Facebook's own Community Standards. Section II discusses the board's structure and membership. Section III addresses the process and criteria for case selection, and Section IV discusses the board's procedures. Finally, Section V considers the board's reach and its power to remedy violations of Facebook's Community Standards.

I. The Need for Governing Principles

If the board is to have independence and authority, its responsibility cannot be merely to interpret Facebook’s internal Community Standards. The reason is that the Community Standards are nearly continuously revised by Facebook’s Content Policy team. Any interpretation of a provision in the Community Standards is hence subject to Facebook reversal, rendering all board decisions potentially advisory.

Instead, Facebook should consider creating or adopting a set of higher-order principles for the board that supersede the Community Standards (and cannot readily be changed or altered by Facebook). These fundamental principles would serve as essential values for Facebook and the board and may guide decisions to revise the Community Standards.

As part of the public input process following the issuance of the Draft Charter, commentators have suggested different potential sources for these higher principles. Some have suggested that the board adopt international human rights law, for example, or free speech law derived from various national constitutions. Alternatively, the board or Facebook could draft a set of original “Guiding Principles,” relying on Facebook’s Mission Statement or other higher order principles from which the Community Standards are derived. These principles would express the competing values that govern both user behavior and Facebook policy on the platform: free expression, safety, privacy, community, and the like.

Although we take no position on the appropriate source of authority (whether a “Constitution” for Facebook or international human rights law), we believe an external source of governing principles is essential to the board’s ongoing legitimacy. Without such governing principles, the board’s decisions will have little enduring weight, as Facebook can override them or render them irrelevant simply by revising the Community Standards. Over time, the board’s own precedents will add nuance to the principles and develop a gravitational force of their own.

II. Structure and Membership

The legitimacy of a social media content review board will depend on its independence and the representativeness of its membership. Like a court or regulatory agency, a truly independent board must be “immune from undue influences from public, private, or other special interest.”⁶ There must be no “inappropriate or unwarranted interference” by another entity,⁷ such as Facebook, in the board’s operations, composition, and decision-making. The structure of the board and its relationship to the platform it oversees – from its inception and selection process through its execution – greatly affect how independent and, thus, legitimate the board will be.

We recommend several features to preserve the board’s independence, including that Facebook delegate board member selection power to a selection committee and fund the board through an independent trust. Measures such as these help ensure that the board serves as an independent check on power, rather than a dependent functionary of Facebook itself. At the same time, effective oversight for a global platform must come from a globally representative body that reflects, in some proportionality, Facebook’s user population. The board’s legitimacy depends not only on the geographic diversity of its members, but also on diversity of expertise, viewpoint, race, religion, gender, and more.

A. Selection Process for Board Members

1. Independent Selection Committee

In a letter to Mark Zuckerberg, David Kaye, UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, stressed that achieving an independent board depends on establishing a “well-entrenched independent mechanism . . . responsible for the appointment, promotion, transfer and dismissal of judges.”⁸ The failure of Google’s recent AI ethics board, including its subsequent dismantling due to internal opposition to the selected members,⁹ demonstrates just how critical a fair, independent, and transparent selection process is.

6 Stanford Global Digital Policy Incubator, Article 19, and United Nations Special Rapporteur on Freedom of Opinion and Expression, *Conference Report: Social Media Councils - From Concept to Reality (February 1-2, 2019)*, 32, <https://perma.cc/6S99-ZZ6K>; see also Kanstantsin Dzehtsiarou and Donal Coffey, “Legitimacy and Independence of International Tribunals: An Analysis of the European Court of Human Rights,” *Hastings International and Comparative Law Review*, no. 37 (2014), <https://perma.cc/K9HF-CA46>; Martino Maggetti, “Legitimacy and Accountability of Independent Regulatory Agencies: A Critical Review,” *Living Reviews in Democracy* (2010), <https://perma.cc/X62R-2AQR>.

7 United Nations Office of the High Commissioner for Human Rights, *Basic Principles on the Independence of the Judiciary*, adopted September 6, 1985, <https://perma.cc/9P5N-BLDC>. The principles were “formulated principally with professional judges in mind, but they apply equally, as appropriate, to lay judges, where they exist.”

8 David Kaye, *Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, OL OTH 24/2019, May 1, 2019, 5, <https://perma.cc/DCH3-KYET>.

9 Kent Walker, “An external advisory council to help advance the responsible development of AI,” Google Blog, March 26, 2019 (updated April 4, 2019), <https://perma.cc/2R9A-LEDT>; see also Nick Statt, “Google dissolves AI ethics Board just one week after forming it,” *The Verge*, April 4, 2019, <https://perma.cc/W79C-UAYM>.

In its Draft Charter, Facebook proposed to select the initial members of the board based on a public “review of qualifications.”¹⁰ The board members would later appoint their own successors. We urge the company to consider alternative selection processes, discussed below. Although our recommendations may slightly increase the cost and time of selection, we believe these costs are a small trade-off for the overall legitimacy and long-term success that the oversight board would gain.

We recommend that Facebook appoint an external body that will be responsible for selecting the first round of board members. Kaye rightfully notes that “the experience of international human rights law indicates that another option under consideration – delegating initial appointments to a selection committee – may be a more robust guarantee of the board’s independence.”¹¹ Facebook should recruit and publicly call for nominations and applications of independent experts to serve on a reasonably small (e.g., five- to nine-member) selection committee. These individuals would ideally come from civil society and academia, with a track record of working at the intersection of technology, policy, and human rights, and with experience in issues pertaining to social media platforms and content moderation. To the extent possible for such a small group, they must also meet the diversity criteria established for the board itself (i.e., demographic, cultural, and geographic diversity).

Once the members of the selection committee are confirmed, they will be responsible for the initial appointment of board members. Although most board members would be selected directly by the selection committee, Facebook may want to consider requiring a small fraction of board members (e.g., twenty percent) to be selected from among a pool of nominees selected by the Facebook user population. For example, board members could be selected from users who volunteer and go through a nomination process, a mechanism similar to selection for Wikimedia’s Arbitration Committee, which resolves Wikimedia content disputes.¹² Alternatively, users who meet minimal eligibility criteria could be nominated by user “election” — that is, by a majority vote of Facebook users who participate in the election process. While this process would increase the legitimacy of the board, there is a risk of electing the most vocal and potentially extremist users to the board (let alone allowing for potential manipulation of the election process by nation-states and other strategic actors). If an outside process is used to nominate some fraction of the board members, then, the pool of “external” candidates will still need to be approved and vetted by the selection committee.

2. Inclusiveness of Process

Whether selection is done by an independent committee or other means, Facebook should ensure that the selection process is inclusive and built on meaningful consultation with diverse stakeholders. Facebook has made unprecedented efforts in consulting with a diverse array of communities in its outreach efforts in the pre-design period for the board. The company should further intensify its efforts to include in the design process vulnerable and marginalized groups. The company must ensure at each stage of the development of

¹⁰ Facebook, *Draft Charter*, 2; see also Nick Clegg, “Charting a Course for an Oversight Board for Content Decisions,” Facebook Newsroom (blog), <https://perma.cc/3FL5-95WK>.

¹¹ Kaye, *Mandate of the Special Rapporteur*, 4.

¹² Wikipedia, “Wikipedia:Arbitration_Committee,” last edited April 29, 2019, <https://perma.cc/MY2W-KMAB>.

the board to include diverse representation along lines of political ideology, gender and gender identity, sexual orientation, race, ethnicity, religion, disability, and other classifications typically targeted for discrimination. Groups that have been historically at risk of censorship or attacks, such as human rights defenders, should also be included in the design and implementation process.¹³

Meaningful consultation requires including affected communities in the conversation – from the initial design throughout the entire implementation process – by listening to them, and, most importantly, acting upon their insights, concerns, and recommendations. We applaud Facebook for its commitment to stakeholder engagement throughout the institutional design process, and urge the company to incorporate lessons learned from this process into its final decision-making. Established best practices for stakeholder engagement, such as those developed in the context of the United Nations’ Guiding Principles on Business and Human Rights, should serve as a guide.¹⁴

B. Membership Criteria

1. Geographic diversity

Board members should be as diverse as Facebook’s global community. After all, “[o]nline content does not fall neatly within national boundaries. The problem is global, and solutions must address and reflect those global dimensions.”¹⁵ At the same time, different countries and regions face unique challenges from social media. Any oversight board thus must balance the need for consistency across the platform with sensitivity to local circumstances and culture. We therefore recommend a global board, divided into regional subsections, such as:

1. US/Canada
2. Latin America
3. Europe
4. The Middle East and North Africa
5. Sub-Saharan Africa
6. South Asia
7. Asia (excluding South Asia)
8. Oceania/Australia

We recommend allocating seats to each regional subsection in some proportionality to Facebook’s user population, while not allowing any one set of voices to drown out the rest.

¹³ Kaye, *Mandate of the Special Rapporteur*, 5.

¹⁴ United Nations Office of the High Commissioner for Human Rights, “Guiding Principles on Business and Human Rights: Implementing the United Nations ‘Protect, Respect and Remedy’ Framework” (HR/PUB/11/04), June 2011, <https://perma.cc/6ZV5-U4QC>; see also United Nations Human Rights Council, *Resolution 17/4 - Human rights and transnational corporations and other business enterprises*, July 6, 2011, <https://perma.cc/28VF-5UMK>; Shift, Oxfam, and Global Compact Network Netherlands, “Stakeholder engagement,” in *Doing Business with Respect for Human Rights: A Guidance Tool for Companies*, 2016, <https://perma.cc/DJ45-5VE6>.

¹⁵ Stanford Global Digital Policy Incubator, Article 19, and United Nations Special Rapporteur on Freedom of Opinion and Expression, 27.

Content moderation will only be legitimate in users' eyes if done with knowledge and understanding of a case's specific local and cultural context.¹⁶ A global board, however, will not always have the necessary knowledge to understand localized needs and fairly apply Facebook's Community Standards. Thus, in addition to a representative board, we recommend establishing a roster of country and regional experts who would support the global board. These experts can help provide additional context when the board itself lacks the requisite cultural knowledge.

A global board will also need to overcome language barriers. To the extent possible, board members should be multilingual and speak at least one language practiced within their assigned region. We also recommend that Facebook define "working languages" for the board, requiring each board member to be fluent in at least one of them, much as international courts do.¹⁷

After the initial selection of board members, specific diversity requirements can be adjusted on a yearly basis by accepting more or fewer members from each region than were initially recommended. Flexibility is necessary, so that the board can adapt if it becomes clear that certain regions have become under- or over-represented.

2. Demographic and Ideological Diversity

The democratic legitimacy of a judicial body directly depends on its ability to reflect the wider society that it serves.¹⁸ Facebook itself has recognized that "the 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."¹⁹

Achieving genuine diversity is one of the most difficult undertakings Facebook will face in creating this board. The board's jurisdiction will cover virtually every country, each of which exponentially expands the number of relevant languages, cultures, ideologies, and experiences represented on the platform. As one commentator remarked, a forty-member oversight board "cannot hope to represent even a meaningful fraction" of the views represented by Facebook's global membership.²⁰ This limitation is a product of Facebook's unparalleled scale and must be publicly addressed by Facebook at the creation of this board.

We make several suggestions to guide the initial board selection committee when considering demographic diversity among board members. First, a reasonable level of gender parity is an achievable metric, which should certainly guide the selection of the initial cohort. Second, when considering other axes of demographic diversity, such as race, ethnicity, or religion, a good starting point may be to ensure regional representation in proportion

¹⁶ *Ibid.*, 16-20.

¹⁷ The International Criminal Court, for example, requires its judges "have an excellent knowledge of and be fluent in at least one of the working languages of the Court." International Criminal Court, *Rome Statute of the International Criminal Court*, Art. 36.3(c), July 17, 1998, <https://perma.cc/C5PW-8SFD>.

¹⁸ Some empirical evidence suggests that there may be substantive differences in decision-making arising from a diverse judiciary. See Rosemary Hunter, "More than Just a Different Face? Judicial Diversity and Decision-making," *Current Legal Problems* 68, no. 1, 2015, 119-141, <https://perma.cc/7RAQ-7MW5>.

¹⁹ Archibong, "Strengthening Safety."

²⁰ Evelyn Douek, "Facebook's 'Oversight Board': Move Fast with Stable Infrastructure and Humility," *North Carolina Journal of Law and Technology* 21, no. 1 (2019), <https://perma.cc/38P2-UT9X>.

to Facebook’s user population, as described above. While this is certainly an imperfect proxy and it will be impossible to represent all constituencies in the user population, it is an important first step to legitimize this body as a truly global oversight board.

Finally, as emphasized above, we consider ideological diversity to constitute a critical component to ensuring the board’s legitimacy. Different democracies strike the balance between free speech and other values, such as security, privacy and anti-discrimination, in different ways. It would be wrong and harmful to the credibility of the board, if, for example, all board members adhered to a single country’s perspective on how to strike that difficult balance. At the same time, Facebook should not feel compelled to provide board representation to authoritarians, merely because many of its users live under such regimes. Nor should it feel compelled to provide representation to those who espouse ideologies completely at odds with the company’s values of free expression, tolerance, equal treatment, and user safety.

3. Professional Diversity

Facebook’s own Draft Charter recommends that members have deep expertise in a range of topics, including “content, privacy, free expression, human rights, journalism, civil rights, safety and other relevant disciplines.”²¹ We agree.

Advisory boards and courts commonly require expertise from their members. Judicial bodies, like the International Criminal Court, generally require that judges have established competence in relevant areas of law.²² Meanwhile, other international bodies, such as ICANN or the Wikimedia Arbitration committee, look for members with technological expertise.²³

Social media oversight boards, generally, and the Facebook board, specifically, should also include members with diverse expertise, drawing from civil society, academia, and industry. Some board members should be legal experts—including a combination of experts in local, national, and international laws and norms. In addition to lawyers, board members should include professionals working in business, media, technology, and nonprofits, as well as community leaders.

4. Eligibility and Independence of Board Members

As important as the board’s own independence is, the independence of individual board members is also paramount to ensure they are not swayed by external interests. We thus propose adding several eligibility requirements to preserve individual board members’ independence.

²¹ Facebook, *Draft Charter*, 1.

²² *Rome Statute*, Art. 36(a); see also International Court of Justice, *Statute of the International Court of Justice*, Art. 2, October 24, 1945, <https://perma.cc/G265-HELL>; Inter-American Commission on Human Rights, *American Convention on Human Rights*, Art. 42, November 22, 1969, <https://perma.cc/PDA9-JJMY>; Organization of African Unity, *African Charter on Human and Peoples’ Rights*, Art. 31.1 & 32, October 21, 1986, <https://perma.cc/9WN9-VLXB>.

²³ ICANN, *Board Governance Guidelines*, “Board Membership Criteria,” amended July 18, 2018, <https://perma.cc/LE69-8C72>; Wikipedia, “Wikipedia:Arbitration_Committee.”

In order to avoid state capture of the board, members of the legislative and executive branches of government should be precluded from serving as board members unless they are retired and unlikely to return to government service. Former judges should be eligible, as long as they do not simultaneously sit on the bench in their respective jurisdictions. Likewise, staff members from international organizations should be considered so long as they do not hold political positions requiring them to represent their country's interests.

Of course, current or recent Facebook employees could not serve on the board if it is to maintain perceived and actual independence. Yet we also recognize that service by former employees could be valuable, given their technical expertise or familiarity with internal Facebook processes. To balance these competing considerations, board members should neither have been employed at Facebook within the last five years, nor should they be permitted to rejoin the company for five years after the end of their term of service on the board.

C. Structure of the Board

1. Size of the Board

The optimal size of an effective social media oversight board depends on a number of factors. The required diversity of its membership is one constraining factor, as only boards of a certain size will be able to provide adequate representation from each region or demographic group. Moreover, the size of the board should be optimized for the expected workload of each member, which will depend at least in part on whether board members serve full time or part time. At the same time, a board that is too big will be unwieldy, costly, and less efficient.

Facebook's Draft Charter proposed a 40-member, part-time board to oversee global content moderation decisions.²⁴ Even with most decisions made by three-member panels (as we recommend below), given the size of Facebook's user population, in conjunction with the huge volume of appealed content — in the first quarter of 2019, for example, users appealed more than 2 million times just for content removals based on the "Adult Nudity and Sexual Activity" sections of the Community Standards²⁵ — we believe this proposed size is too constraining to allow Facebook to achieve a truly diverse board able to represent its global user base. If board members are to serve part-time, the board must be significantly larger to handle any meaningful volume of appeals. Alternatively, board members could serve full time, although this might limit Facebook's ability to attract a high-profile slate of candidates from diverse fields.

Finally, given the large size of the board, a Chair of the Board will be critical to managing internal processes and ensuring that the board fulfills its duty as an independent check on Facebook. In addition, each regional subsection should have a vice president to handle matters of particular importance to the region.

²⁴ Facebook, *Draft Charter*, 1.

²⁵ Facebook, *Facebook Transparency Report: Community Standards Enforcement Report*, May 2019, <https://perma.cc/YC4V-J6JD>.

2. Board Term

We recommend that board members serve staggered, multi-year terms (e.g., three- to five-year terms) and be subject to removal only for cause.

Multi-year board terms are critical to effective oversight, and are typical for analogous institutions. For example, members of the ICANN Board serve three-year terms,²⁶ and judges at the International Court of Justice serve nine-year terms.²⁷ This structure enables board members to serve long enough to cement institutional memory and provide some measure of continuity. However, given the importance of engaging diverse viewpoints, we suggest allowing each member to serve only one term. Term limits ensure that the initial cohort of board members will not come to dominate the board and become entrenched. Turnover in board membership is necessary to ensure responsiveness to changing dynamics in the social media environment and to protect the board against any individual member having too much power.

A 3-to-5-year term limit requires the board to implement additional measures. First, it is important to have staggered terms, not only to ensure continuity of institutional knowledge, but also to allow newcomers ample opportunity to contribute to the board.²⁸ Second, board members should not have absolute immunity for misconduct, such as for a breach of board confidences, and there should be mechanisms for holding them individually responsible and removing them, if necessary. It is common in other systems of governance that an external body votes on removal; for example, U.S. federal judges can be impeached and removed by Congress.²⁹ However, allowing Facebook, or even the selection committee, to remove board members at their whim could pose serious problems for the board's independence and legitimacy. Thus, serious thought should be given to which entity (the board as a whole, the standing selection committee) has the power of removal, under what conditions, and according to what kind of process. Removal provides a potential backstop on abuses of power by any individual member, but should be deployed only for good cause.

3. Staff

We agree with Facebook that the board must “be supported by a full-time staff.”³⁰ The staff should be employed by the board, not Facebook, and should include: 1) subject-matter experts on particular regions, content types, and/or technologies; 2) staff to recommend cases for board review; and 3) general staff members to assist the board members in their duties. These permanent staff members would be complemented by experts or others on an ad hoc basis, as required for particular cases.

With a global board, it is impossible to have members with expertise on all topics that might arise in individual cases. Thus, the board will require a group of subject-matter experts who are familiar with major issues and concepts from around the world. Each of these staff members could be responsible for a particular region, maintaining contacts with other experts in their region and bringing them in to offer additional expertise when necessary. Additionally, this standing staff should include experts in other domains, including technical ones,

²⁶ ICANN, *Board Governance Guidelines*.

²⁷ International Court of Justice, *Statute of the International Court of Justice*, Art. 13.

²⁸ The ICJ, for example, has staggered terms. *Ibid.*

²⁹ Under the United States Constitution, the House of Representatives files articles of impeachment and the Senate conducts the impeachment trial. U.S. Constitution, Art. I. For an explanation of the judicial impeachment process in the United States, see Douglas Keith, Brennan Center for Justice, “Impeachment and Removal of Judges: An Explainer,” March 23, 2018, <https://perma.cc/8HLL-TV24>.

³⁰ Facebook, *Draft Charter*, 1.

because not everyone on the board will have the requisite knowledge of the technical implications of their decisions or the practicality of implementing those decisions. This structure could be loosely based on the U.S. Congressional Research Service, which offers nonpartisan research and report service to Members of Congress.³¹

Additionally, a single board member's workload is likely to be too much to handle by that member alone, especially if members work only part-time. We therefore recommend that each member be given the resources to hire at least one full-time staff member. This staff member would be in charge of writing opinions or performing research on behalf of the board member, thus lessening that member's workload and allowing that member to focus his or her attention on hearing and deciding cases.

4. Funding

Finally, the source and dynamics of the board's funding will be critical to maintaining its independence. We considered three possible funding sources: Facebook itself, international government entities, and nonprofit donations. International, government-backed funding could aid in ensuring the adoption of the board as a truly global entity. However, the risk of state capture dramatically increases with the introduction of large government actors into the funding model. Third-party nonprofits might also attempt to exert undue influence on the board, as any external entity capable of contributing a significant financial donation may be associated with a particular agenda.

We find funding through Facebook to be the most viable option. This option is practical: Facebook has the resources to pay for it. We also believe it is Facebook's responsibility to bear the costs of enforcing its own Community Standards and ensuring fairness in content moderation. Facebook should consider a funding mechanism to pledge a fixed budget for an extended period of time. Facebook could do so by setting up an independent trust, controlled by people with no relation to Facebook, which actually controls disbursement of funding to the board.

31 Congressional Research Service, *About CRS*, April 16, 2019, <https://perma.cc/6MWK-N8VY>.

III. Case Selection

To maintain legitimacy as a body independent from Facebook, the board must be able to select its own cases from among the pool of eligible content moderation decisions. To achieve procedural fairness, any interested party who has exhausted his or her internal appeals and wants to appeal a content decision to the board should be able to petition for review. This includes posters of content, flaggers of content, targets of content, and interested third parties, such as Facebook itself, non-governmental organizations (NGOs), other companies, and governments. The petition forms should be structured, requiring petitioners to expend some effort and eliciting relevant contextual information to guide the board through selection and decision.

To cope with the vast scale of disputed content decisions, we propose a three-tiered system composed of Elevation Teams, Selection Panels, and Decision Panels. A set of Elevation Teams, organized by region and content category and comprised of full-time staff members, sorts through the large pool of petitions to present a subset to a Selection Panel for consideration. The Selection Panel, composed of rotating board members, considers the cases nominated by the Elevation Teams, alongside a few cases picked as a random sample from among all content takedowns, and decides which cases to select for review by future Decision Panels.

When choosing which petitions to elevate and ultimately select, both the Elevation Teams and Selection Panels should follow four guiding criteria, prioritizing cases that:

- Raise important undecided issues,
- Reveal considerable variance between moderators,
- Involve potential miscarriages of justice, or
- Conflict with prior board precedent and/or governing principles.

In order to ensure that no single type of petition dominates the board's decisions, both sets of case selectors should also consider the following types of diversity: geographic region, petitioner type, and violation category.

Given the vast quantities of contested content removed by Facebook and appealed by users, Facebook may need to modify its existing internal appeals system for the board's selection process to be effective. First, to apply the guiding criteria, selectors need to be able to access and sort petitions based on metrics such as how often particular pieces of content are shared, flagged, and removed, and what rules Facebook content moderators followed to make decisions on those pieces of content. Second, Facebook should ensure that adequate appeals mechanisms are in place for all types of content and that the board has authority to review those decisions.

A. Cases Eligible for Board Review

What types of cases should even be eligible for board review? Facebook has stated that the board's primary function will be to review specific content decisions made when enforcing the Community Standards.³² We offer a few caveats.

Facebook asserts that the board should “not decide cases where reversing Facebook’s decision would violate the law.”³³ After all, since Facebook is a public company, it cannot be compelled by an external body to act outside the law and thus put its members in legal jeopardy. That said, some alleged violations of national law are in fact pretextual. For example, it has been reported that freedom of speech violations have been committed under the guise of intellectual property violations.³⁴ In particular, copyright claims can be deployed by governments to limit political dissident speech, interfere with elections, or just threaten censorship in general.³⁵ Excluding these cases from board review would raise serious concerns and undermine the board’s ability to check attempts at chilling speech on the platform.

In addition to content takedowns, some account takedowns should be within the board’s purview. Accounts are sometimes shut down due to content standard violations (and not for other policy reasons, such as impersonation or fake names). We think it is important for account disabling to be eligible for review by the board since the effect of disabling an account flows from a Community Standards violation and chills speech.

Finally, while a danger lies in writing off any category of content from appealability, it may be necessary for spam, due to the sheer volume of potential appeals. From the data Facebook has published for the first three months of 2019, the volume of spam content pieces at issue dwarfs all other published categories combined. Spam alone accounted for 1.8 billion pieces of content removed, 20.8 million appealed, and 15.1 million appeals denied. The equivalent figures for all non-spam categories published combined, per three-month period, are: 73.0 million removed, 4.1 million appealed, and 3.4 million appeals denied and therefore theoretically eligible for appeal.³⁶ At the very least, if spam decisions can be appealed, review by selectors would need to be highly abbreviated.

32 Facebook, *Draft Charter*, 3.

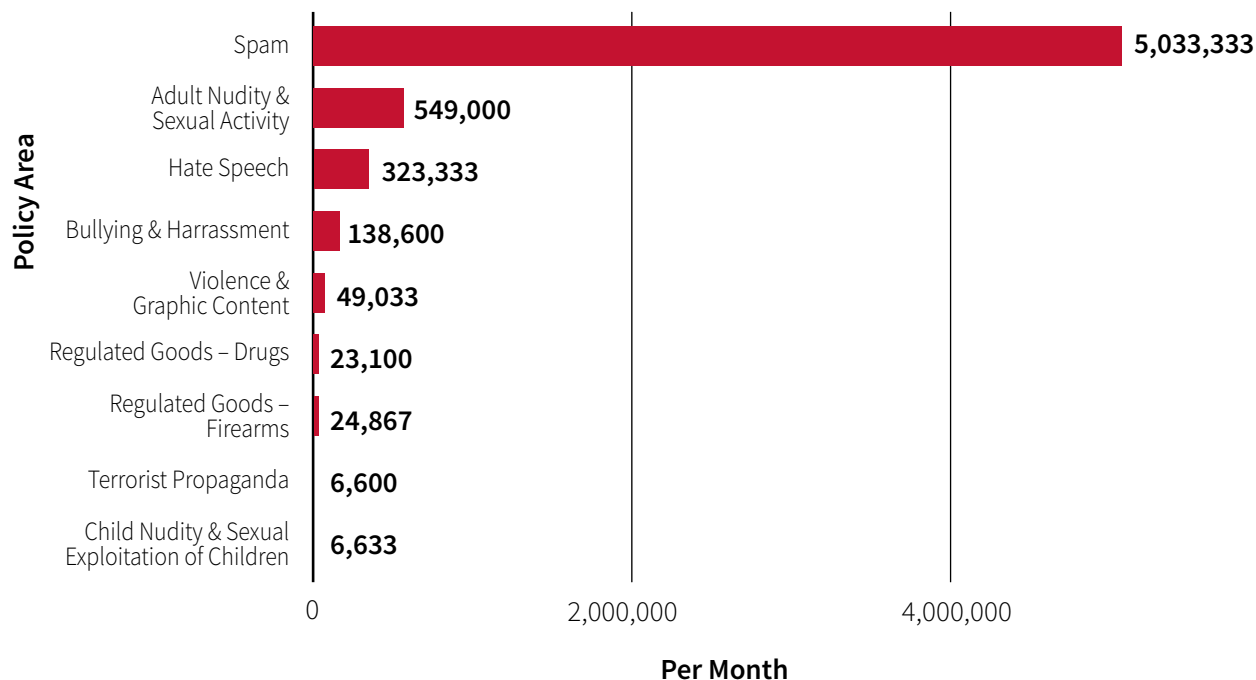
33 *Ibid.*

34 See Daphne Keller, “Who Do You Sue? State and Platform Hybrid Power Over Online Speech,” Aegis Series Paper No. 1902, January 2019, Hoover Institution. <https://perma.cc/K5TH-4SHD>; Irene Caselli, “Taking Down the Critics: Is Ecuador’s Government Silencing Social Media Users?” *Index on Censorship* 43, no. 3 (September 2014), <https://perma.cc/8CDG-Z4AS>.

35 See Corynne McSherry, Electronic Frontier Foundation, “Platform Censorship: Lessons From the Copyright Wars,” September 26, 2018, <https://perma.cc/N4YJ-MTYM>; Stephen Gray, “Copyright Strikes as a Tool for Censorship, Bullying, Extortion,” *NYU Journal Of International Law And Politics* (blog), March 7, 2019, <https://perma.cc/2AUX-UA9K>; Alex Hern, “Revealed: How copyright law is being misused to remove material from the internet,” *The Guardian*, May 23, 2016, <https://perma.cc/TD59-JRHP>.

36 Facebook, *Community Standards Enforcement Report*. Note that, according to the guide accompanying the report, Facebook’s enforcement statistics do not incorporate second-level appeals or appeals by content reporters. See Facebook, “Understanding the Community Standards Enforcement Report,” May 2019, <https://perma.cc/8GTH-TFRF>.

FIRST-LEVEL POSTER APPEALS DENIED BY FACEBOOK



Calculations based on Facebook’s Community Enforcement Report, May 2019.

B. Petitioners

To promote the accessibility, independence, and accountability of the board, we recommend implementing an open petition process. This petition process should be accessible to (1) individuals who are unsatisfied with the result of their appeal via the Facebook process; (2) organizations (i.e. NGOs, companies, and governments) that disagree with a content moderation decision taken on their own content or the content of an individual; and (3) Facebook itself. Petitions to the board should be tailored to a content takedown decision and can additionally call on the board to specifically reconsider a policy detailed in the Facebook Community Standards. We also recommend the inclusion of a defined number of cases selected at random from the pool of content takedowns that were not appealed, in order to acquaint the board with the operation of the Community Standards in garden variety, as opposed to edge or hard cases, and to allow them to provide oversight even in run-of-the-mill cases.

We recognize that opening the petition process in this manner might open the floodgates and require a substantial investment of resources. However, we believe this accessibility is essential to the legitimacy of the board. By opening access to the board to the Facebook community at large, and by empowering the board to select from the petitions it receives, the board’s independence and accountability to the broader Facebook community can be secured.

1. Who Can Petition the Board?

a. Original Parties

The first category of petitioners that can directly petition the board are individuals that have (1) posted content that was removed by Facebook, (2) flagged content to be taken down, or (3) been victimized or targeted by content that has been posted on Facebook.

For content that is appealable, individuals generally must have exhausted the internal appeals process prior to submitting a petition to the board. That said, our understanding of the current internal appeals system is that only individuals who have viewed content are able to petition for review. The exhaustion requirement thus may need to be relaxed or the internal petition process expanded in instances where parties whose interests are clearly implicated (e.g., a user who is depicted in but not exposed to content) are unable to exhaust internal mechanisms.

b. Outside Groups

Outside groups such as NGOs, companies, and governments should also be able to submit petitions directly to the board. These outside groups can petition regarding the takedown of their own content (e.g., if they have a Facebook Page that they use to post to their communities). Facebook should also consider allowing these outside groups to petition on behalf of individuals, at least where individuals cannot necessarily petition the board themselves (e.g., where the victims of a post are unidentifiable). Governments will not be able to petition the board regarding compliance with their national laws.

To address concerns that well-resourced outside groups will be better equipped to prepare petitions to the board (and thus be selected over individual petitions), we recommend that case selection targets be set for the various types of petitioners. These targets will be detailed later in this section and will function to ensure an equitable distribution of the board's resources.

c. Facebook

Because of Facebook's unique perspective and access to information about potential cases, the company should be able to petition the board directly. We envision several different sources of Facebook petitions: (1) nominations by content moderators, (2) contentious content seeing a high volume of flags or appeals from users, and (3) escalations by Facebook employees. Facebook petitions to the board should not be granted automatically and should be considered alongside the petitions from individuals and outside organizations.

We also considered whether or not there should be a "crisis channel" for emergency decisions. This crisis channel would bypass the full case selection process and funnel a case directly to a Decision Panel for review. As a recent example of a possible crisis event, consider the Christchurch, New Zealand shooting in which the perpetrator of the attack live streamed his actions on Facebook.³⁷

A crisis channel would have the benefit of allowing Facebook to call on an independent third party, the board, to reach an independent decision on the content while the crisis is ongoing. Given the sensitivity, reach, and potential safety implications of crisis content, an impartial board imbued with expertise would be particularly beneficial.

³⁷ Cade Metz and Adam Satariano, "Facebook Restricts Live Streaming After New Zealand Shooting," *New York Times*, May 14, 2019, <https://perma.cc/78SS-2KCR>.

There are, however, some disadvantages to implementing a crisis channel. First, given the time-sensitive nature of these issues, some board members would have to be on call throughout the year to monitor this channel. Second, there is the potential for abuse, especially without a clear definition of what qualifies as a “crisis.” Too many escalations via this channel would take resources away from individual petitioners and outside organizations. If Facebook does adopt a crisis channel, crisis escalations should be capped (although setting that numerical cap would almost certainly be arbitrary).

An alternative to creating a crisis channel is to keep the status quo. Facebook would continue to be responsible for making time-sensitive content moderation decisions in times of crisis, and eligible petitioners would raise petitions with the board after the fact if they disagreed with the action taken by Facebook.

d. Random Sample

Finally, we suggest that the board consider an allotted number of randomly generated cases per month, to perform a quality assurance function on the content moderation process, to ensure that the board is not only seeing the edge cases, and to compensate for any inequities in the petitioning process. These randomly selected cases should be pulled from: (1) the pool of non-exhausted Facebook appeals and (2) the pool of content takedowns, such as intellectual property disputes, for which internal appeals are not available.³⁸

Including randomly selected cases will ensure that the board is educated on and aware of decisions being made in the run-of-the-mill cases, rather than only having a view into the edge cases. Random review will serve two purposes. First, it will ensure fairness. As we understand it, not all users are necessarily aware of the various appeal offerings, nor do all users have the resources (education, language ability, etc.) to go through the appeals process. Including a random sample of these cases will ensure that these members of the community are also heard by the board. Second, it will serve as a quality control process. The board will be looking to ensure that content was appropriately tagged for violating a particular Community Standard and that the takedown was justified.

Cases generated by random sample will skip the first stage of the case selection process, but Selection Panels will still review for frivolous cases and compare with the wider set of potential cases before passing these cases on to Decision Panels.³⁹

2. How to Petition

Petitioners will be able to file petitions related to content that was removed or left up during the Facebook content moderation process. (This will include content takedowns that resulted in an account or Facebook Page being disabled.) Petitioners should file petitions to the board through topic-specific forms. For example, if a petitioner’s content was removed because it violated the Hate Speech Community Standard, the petitioner would fill out the Hate Speech form. As part of their petition, petitioners should have the option to request a change or clarification of the Facebook Community Standards.

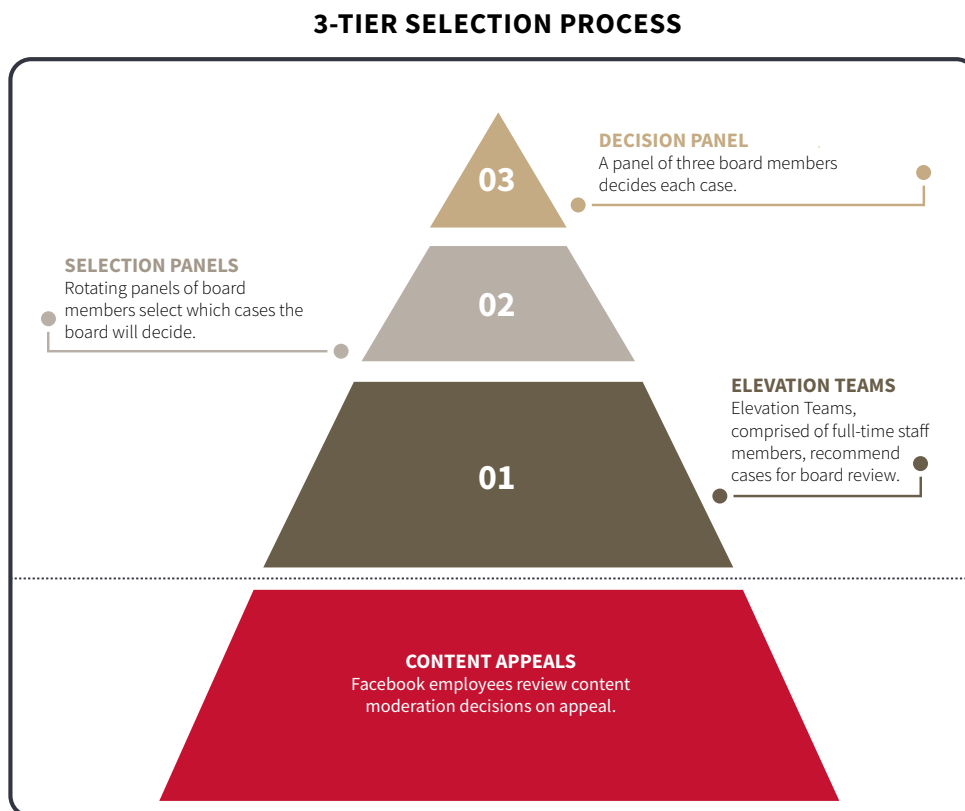
Petition forms should be structured and detailed to ensure that the board has the information and context they need to make informed decisions, as well as to create some minimal hurdle to reduce the overall volume of petitions. An example form is included in the Appendix.

³⁸ Facebook, *Community Standards Enforcement Report*.

³⁹ Subsection IV.B elaborates on these procedures.

C. Selection Process

Given the challenges posed by the large volume of appeals that are submitted every day, we propose a three-tiered selection process. In the first tier, appeals to the board are reviewed by an *Elevation Team* which will be comprised of permanent staff members. Members of the Elevation Team select some number of cases to be reviewed by the board. In the second tier, a rotating panel of board members (a *Selection Panel*) will choose a subset of the cases recommended by the Elevation Team to present to the larger board for decision. Finally, a *Decision Panel* comprising a portion of the board members will deliberate on all cases recommended by the Selection Panel, and ultimately reach a decision on each case. Each tier of this proposed selection process is described in greater detail in this subsection.



1. Tiers of the Selection Process

a. Elevation Teams

Elevation Teams will recommend cases for board review. Elevation Teams should be able to join similar cases together at this stage, to ensure the board has full insight into a given issue. For example, if multiple petitions concern the same piece of content or related pieces of content, they should be joined together before being sent to the Selection Panel.

We believe that Elevation Teams should be comprised of full-time staff members who report to the Chair of the Board and whose work is overseen by a board committee. In order to ensure fair consideration of appeals across different content categories and geographic regions, the staff members should collectively have content-level expertise across all appealable categories, and proportionately represent the eight core geographic regions. This

can be achieved in two ways: Teams can be divided by region with content-level experts on each regional team, or teams can be divided by content type with representatives from each region on every team.

Organizing teams by region will arguably allow for greater range of language capabilities, as well as more effective consideration of region and culture-specific appeals, particularly when local or regional context is important to the appeal. For example, this would be especially true when judging whether or not a piece of content is “newsworthy” in a particular region. Having content-level experts on each regional team (e.g., experts on hate speech, bullying, violence and graphic content, etc.) should allow the team to collectively make nuanced decisions on different types of content. However, compared to an organizational model where Elevation Teams are organized first by category and then by region, this regional model makes it somewhat more difficult for Elevation Team members to choose cases representing the most significant undecided issues within each category type to bring to the Selection Panel.

Organizing Elevation Teams by category is advantageous in that it enables each team to gain nuanced content-level expertise. For example, an Elevation Team focused on hate speech would review all appeals of content taken down for a hate speech violation, and would be able to compare the relative significance of different hate speech appeals, selecting those that are more controversial or broadly representative of wider trends. While each content-focused Elevation Team would be comprised of staff members from every region, it is clear that these teams would be unable to integrate region-specific understandings to the same extent as regional elevation teams.

Since the primary purpose of the Elevation Teams is to filter through appeals to present to the board, we believe that members of the Elevation Teams should be chosen by and accountable to the board itself. We view the Elevation Teams as separate from Facebook’s content moderator teams: That is, Elevation Team members should not also be content moderators, and they should work independently from Facebook. Having board members each select a number of staff members to work on the Elevation Teams would help ensure geographic and professional diversity, mirroring the diversity of the board itself.

The required number of Elevation Team staff members will depend heavily on the number of petitions the board receives. We estimate, however, the potential need for more than a hundred Elevation Team members. Let us assume that only 20 percent⁴⁰ of users appeal their final determination by Facebook—likely an underestimate, unless Facebook introduces significant friction into the appeals process. Even with that conservative assumption, 60 full-time staff members reviewing petitions for 35 hours per week would have to sort through 27 petitions an hour to keep up with our estimated number of incoming non-spam petitions. To achieve a still aggressive but more manageable rate of 10 petitions per person per hour, the Elevation Teams would need over 150 full-time staff members.⁴¹

Moreover, the number of cases that the Elevation Team recommends to the Selection Panel will vary with the total volume of appeals and the size of the board itself. If the board had 60 members for example—already more than Facebook is recommending—the Elevation Teams might collectively send no more than 90 pieces of content to each twice-monthly Selection Panel, which in turn would select no more than 20 pieces of content to be heard by Decision Panels.⁴²

b. Selection Panel

Cases would then go to a Selection Panel, a rotating subset of board members who decide which cases should be brought to the entire oversight board. Members of the Selection Panel will be board members who rotate onto the Selection Panel on a monthly or bi-monthly basis. We propose a Selection Panel of three members, who represent different regions and have expertise in different subject areas. In that way, the Selection Panel is not unfairly biased towards any one region or subject area for a particular month. The rotational aspect of the Selection Panel will also help ensure that the panel brings a diverse set of cases to the overall board.

The Selection Panel should forward a diverse set of cases on to Decision Panels. First, a small percentage of cases should come from the random sample, as opposed to from Elevation Teams (e.g., 10 percent of all cases forwarded to the Decision Panel). Second, the selection of all cases should be subject to minimum target percentages by geographic region, petitioner type, and violation category. We discuss these targets in more detail in the next subsection.

40 See, e.g., Theodore Eisenberg, “Appeal Rates and Outcomes in Tried and Nontried Cases: Further Exploration of Anti-Plaintiff Appellate Outcomes,” *Journal of Empirical Legal Studies* 1, no. 3 (2004), <https://perma.cc/SV5P-H3K3> (noting that 21 percent of cases with a definitive judgment for the plaintiff or defendant are appealed in the U.S. federal court system).

41 These estimates are based on Facebook’s *Community Standards Enforcement Report* from May 2019. For January through March 2019, across all violation categories except for spam, Facebook decided to keep down 3,363,500 pieces of content that were appealed. Dividing this by three gives a per-month estimate of 1,121,167. Given a volume of 1,121,167 pieces of non-spam content eligible for appeal per month, and assuming only 20 percent are actually appealed to the board, we estimate that Elevation Teams would need to sift through 224,233 petitions each month. Given that volume of petitions, 150 Elevation Team members spending 35 hours per week would thus each have to go through an average of 11 appeals per hour. Since Facebook’s published appeals data does not include second-level reviews, this number may fluctuate downward, but since this data also does not include reporter appeals and unappealable categories of content, the number could also conceivably fluctuate upwards.

42 The case numbers here represent the maximum that a 60-person part-time board could reasonably consider. Given an estimate of each case requiring 8 hours of member attention and a member being able to spend half of his or her time on the case itself (and the other half on case selection, board administration, and other tasks), 60 board members spending 8 hours per week and sitting on 3-judge decision panels could hear a maximum of 40 cases per month. That caseload would equal just 0.02% of the total pool (224k).

The Selection Panel’s choice whether or not to refer a case to the Decision Panel should be public to the extent possible. For non-referrals, at least the geographic region, petitioner type, and violation category should be published. For referrals, the panel should also publish which of the four guiding criteria provided the basis or bases for the selection decision.

c. Decision Panels

The Decision Panels will be comprised of three board members not currently serving on the Selection Panel. A Decision Panel will not participate in the selection of its own cases, instead hearing all cases referred to it by the Selection Panel, in order to ensure adequate diversity and objectivity in its decision-making process. Where appropriate, a Decision Panel may elect to dispose of certain cases on a summary judgment basis—that is, without the presentation of additional evidence—rather than requiring full review.⁴³

2. Basis for Selection

a. Distributive Considerations

To protect the board’s legitimacy, all judges and staff considering which cases to elevate and select should consider three key dimensions of representativeness: geographic region, petitioner type, and violation category.

Geographic Region:

Case selectors should aim to select cases in rough proportion to the distribution of users across different regions of the world. If the majority of cases heard by the board arose in any one country or continent, that would fatally compromise the global legitimacy of the board. On the other hand, adhering to requirements to select numbers of cases strictly proportionate to the user populations of each country would be unworkable and hamstringing the case selection process. Measuring geographical diversity at the level of regions or continents would provide more flexibility and workability to selectors.

Selectors could consider one or both of two types of user population measures: (1) the distribution of platform users across different regions of the world and (2) the distribution of petitioners across different regions of the world. Prioritizing the distribution of users may further perceived legitimacy, but prioritizing the distribution of petitioners may better serve procedural fairness. At the same time, focusing on the distribution of petitioners may over-sample certain populations with better information about and access to appellate rights and exacerbate structural bias against petitioners from under-served linguistic groups. Accordingly, we recommend aiming for regional diversity in proportion to user population, rather than petitioner population.

Petitioner Type:

Case selectors should ensure that minimum proportions of cases raised by individual content posters, flaggers, and targets reach the board, alongside cases raised by Facebook and other institutions such as non-profits, companies, and governments. To preserve perceptions of procedural fairness and actual legitimacy, no one type of petitioner should dominate the cases heard by the board. We focus here on the petitioner category, including both individuals (posters, flaggers, and targets of content) and non-individuals (non-profit organizations, governments, companies, and Facebook itself). Although diversity of other petitioner characteristics also warrants consideration—no one race, gender, age range, sexual orientation, etc., should dominate the

⁴³ A case decided on a summary judgment basis would bypass the usual board procedures, described in Section IV. Instead, the Decision Panel should issue a short order with its decision.

conversation—the board’s institutional design risks systematically under-sampling certain petitioner types unless they receive programmatic attention. For instance, if the board hears only cases raised by petitioners who want content to be restored—and none by petitioners who want content to be removed—lopsided rulings will result.

Selecting cases without regard to whether or not individual users’ voices are being drowned out by better-funded companies and NGOs would imperil the board’s legitimacy and impair procedural fairness. Adding procedural friction to the petitioning process is essential for meaningfully filtering the vast body of internal appeals down to a scale and into a form that a handful of Elevation Teams can manually review. However, companies, NGOs, and governments will generally have more ability than individuals to navigate lengthy forms. Even if petitioning requires only filling an unstructured text box, better-resourced institutions will have more capacity than would lone individuals to persuasively articulate why the board should take their case. Accordingly, requiring some proportion of cases presented to come from individual users who posted or flagged content—or are targeted by the disputed content—is necessary. Otherwise, operating via the path of least resistance, selection bias may shut these individuals out.

Equally importantly, if Facebook itself dominates the board’s docket, that would fatally undermine the board’s legitimacy as an independent voice. While Facebook will have special insight into what controversies most impact Facebook users, divide Facebook’s content moderators, and deserve emergency escalation, it cannot equally represent the multiplicity of perspectives brought to the table by other individual and institutional actors.

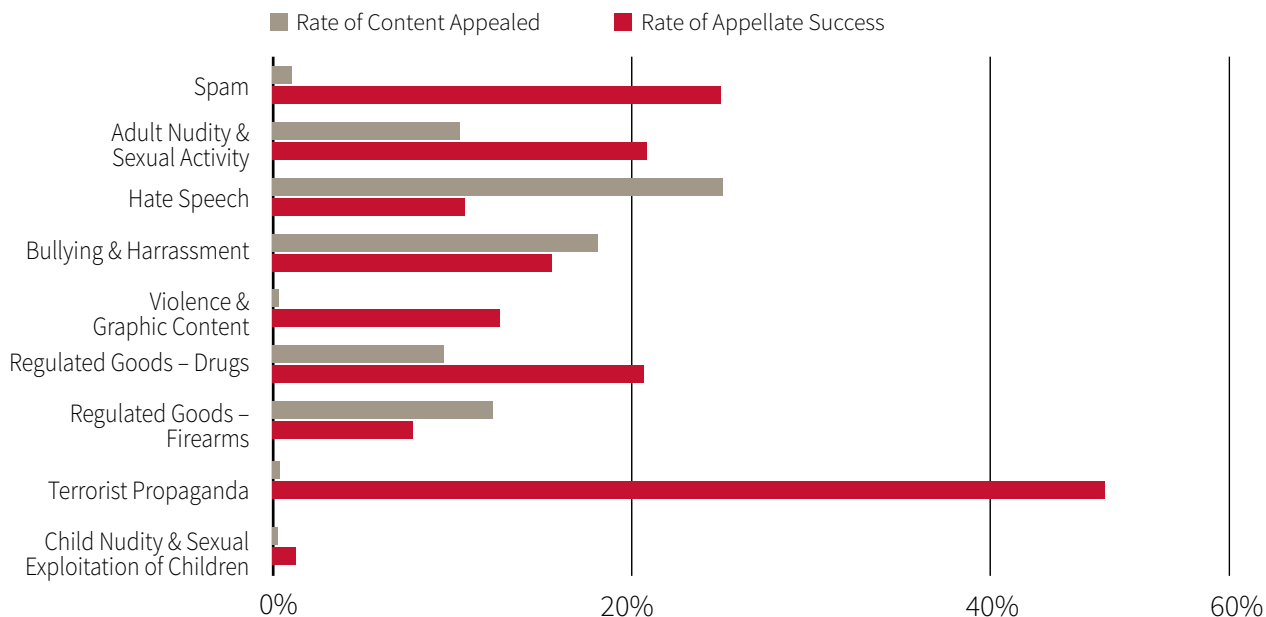
Violation and Sanction Category:

Case selectors should enable the board to hear cases from a broad variety of Community Standards categories rather than focusing narrowly on one or two issue types, such as bullying or adult nudity. Some Community Standards categories will naturally receive more attention from the board than others. For instance, hate speech would fairly account for more cases heard by the board than violence and graphic content. Facebook users appeal more than 27 percent of removals of hate speech, compared to around 0.5 percent for removals of violence and graphic content.⁴⁴ This makes sense as a function of both the complexities of Facebook’s relevant hate speech Community Standards and users’ attachments to expressing what they see as political opinions. However, a board that never hears a graphic content case will not represent procedural fairness for any of the users whose approximately 147,100 internal appeals of graphic content decisions were denied in the first three months of 2019.⁴⁵

44 Calculated using statistics from Facebook’s *Community Standards Enforcement Report*, for January through March 2019.

45 *Ibid.*

FACEBOOK FIRST-LEVEL POSTER APPEALS: RATES



Calculations based on Facebook's Community Enforcement Report, May 2019.

b. Mechanism: Targets, Quotas, or Intangibles

In order to ensure case selector consideration of these dimensions (geography, petitioner type, violation category), specific targets or quotas could guide the Elevation Teams' choices of which cases to present to the Selection Panel or the Selection Panel's choices of which cases to select for review by Decision Panels.

Of course, articulating numerical quotas or targets has some disadvantages. Quotas may reduce the flexibility of selectors to choose the most pressing cases free from other constraints. They also increase the complexity of the selection process, requiring more coordination between Elevation Teams and/or more time dedicated by Selection Panel members to ensuring the math works out.

We believe the benefits outweigh these costs, though. Quotas will ensure that the board actually decides on a diverse set of cases, and they will encourage more coordination between Elevation Teams. Instead of each Elevation Team potentially just picking its top cases in isolation, the teams will have to work together to achieve a mandated configuration.

Given these tradeoffs, we recommend that target percentages operate as minimums for the Elevation Teams but not bind the Selection Panels. Since Elevation Teams have more manpower, are dedicated to board business full-time, and are picking larger numbers of cases at a time, they will be better able to navigate the algebraic logistics of ensuring minimum levels of representation than the few board members sitting on a given month's Selection Panel. To help keep Selection Panels somewhat accountable to these diversity considerations, their decisions to refer or not refer cases to Selection Panels should be made public, or at least information about each case's geographic region, petitioner type, and violation category.

c. Guiding Criteria

Case selectors (both Elevation Teams and Selection Panels) should prioritize cases that (1) raise important undecided issues, (2) reveal considerable variance between moderators, (3) involve potential miscarriages of justice, and (4) conflict with the board’s prior precedent and/or governing principles.

Raise Important Undecided Issues: The significance of a particular undecided issue or set of content should be the most important criteria selectors consider. Clarifying how the Community Standards should be interpreted when their application to a piece of content is unclear is one of the board’s fundamental functions. The analogy to a U.S. judicial framework would be an issue of first impression. This criterion incorporates how often a particular piece of content or substantially similar pieces of content are posted, flagged, and removed. Therefore, access to these metrics will be key for meaningful and efficient case selection.

Reveal Considerable Variance: When the same or substantially similar pieces of content are treated differently by large numbers of moderators, that not only suggests an area where the Community Standards may be ambiguous and in need of clarification, but it also potentially identifies a gap in Facebook’s enforcement policy. The U.S. judicial analogy would be a circuit split, in which lower courts disagree on how to interpret a certain principle. For this criterion, access to how moderators decided to remove or restore disputed content will be important. This would help selectors, as well as the board members ultimately deciding on a case, pinpoint where disagreement arose and tailor their recommendations accordingly.

Potential Miscarriages of Justice: Where a selector feels strongly that an appealed content decision was deeply wrong—even if other selectors feel that it was right—that may point to a close issue that calls for the board’s intervention. This may be particularly likely when the consequences of a content decision are especially severe. This criterion would enable the board to act when members feel that moderators made a grievous error, that a particular piece of content has fundamental importance, or that there should be an exception to a Community Standards rule given its unintended consequences.

Conflict with Precedent and/or Governing Principles: As described in Section V, we believe the board should be able to set its own precedent. Where content moderators take down content inconsistently with past board precedent, board review is warranted. Likewise, selectors should also consider conflict with the board’s own Constitution or “Guiding Principles” (to the extent they exist) as a reason to hear a case.

D. Information Required about Facebook’s Appeals Process

The process of selecting cases requires granular issue-tagging of content takedowns and appeals, data to shed light on the magnitude of certain content takedown petitions, and potentially an expansion of the appeals process.

When a petition is submitted for board review, the case selectors should be able to see what Community Standards category the content violates (or is alleged to violate) at the most granular level possible. Assuming that content moderators are provided with detailed decision trees to guide their decisions,⁴⁶ the board should be able to see which decision tree was applied to the piece of content being petitioned, and—if possible—how

⁴⁶ See The Economist, “The unacknowledged legislators of the online world,” June 15, 2019, <https://perma.cc/29PP-3BJ5>.

it was applied. This categorization will be critical for compliance with category thresholds (to ensure diversity of cases) through the case selection process. The decision trees will also provide needed context for individuals selecting cases to be heard by the board, as well as aiding Decision Panels in their assessment of the cases they hear.

Furthermore, to aid the Elevation Teams and Selection Panels, the board should have access to data indicating how many times a particular piece of content has been flagged or appealed by users. Additionally, the Elevation Teams will need to know which users and pieces of content have satisfied the exhaustion requirement (i.e. have gone through the full Facebook appeals process where available) before petitioning the board.

Finally, we recommend that a full appeals process be available for all content types, if it is not already. Assuming that all exhausted cases would be eligible for board review, an expanded appeals process would mean that petitioners must expend more effort before their case reaches the board. This, in turn, will minimize the volume of petitions that the board receives, helping with the challenges of scale that we anticipate.

IV. Procedures

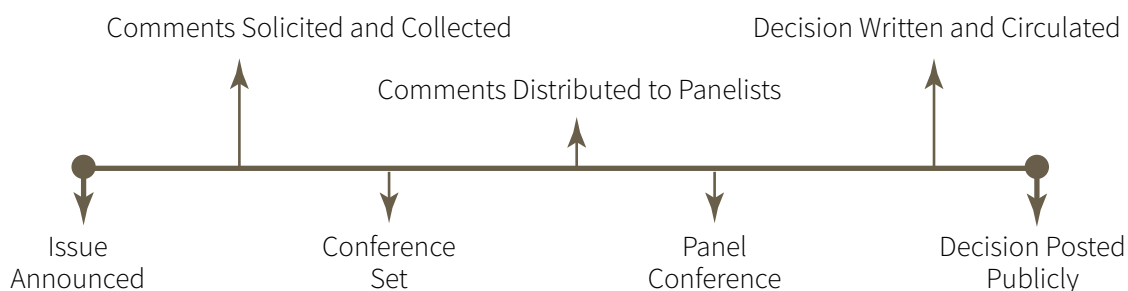
Clear and effective procedures are essential for the legitimacy of the social media oversight board. This Section lays out the process the board should follow once an issue is selected for full board review. First, we outline the general timeline for the board’s process. Second, we explain how the board will announce an issue publicly and solicit comments from individuals directly affected and outside stakeholders, where appropriate. Third, we consider what should be included in the record for the panel, the governing standards and principles that the panel should apply, as well as the makeup of the panels. Finally, we discuss written opinions, procedures for en banc review, and necessary steps to ensure public transparency.

A. Overview: Timeline

The board should resolve issues expeditiously. The board will likely be viewed as more legitimate if it is able to quickly review cases and provide public guidance. Memes, viral stories, and misinformation move fast. A slow response time may undermine the effectiveness of the board’s decisions. Nonetheless, well-researched and reasoned decisions take time.

Facebook should set a timeline by which a three-member panel must make and publish an opinion after an issue is announced. This timeline should allow enough time for deliberation, without letting the issue languish on the board’s case list. While dependent on the scale of the board and its caseload, the time from selection to decision could be targeted to around 30-45 days. The board should publicly announce an issue it has taken under review on its website. After an issue is announced, staff will solicit and collect comments from the original poster, anyone who believes themselves injured by the posting, from user advocate third parties, regional and/or subject matter experts, and from Facebook. This information should be collected expeditiously, while providing enough flexibility in the timing for anyone interested in responding. (The board might consider requiring all comments within, say, 14 days, while extending that period for particularly complicated issues that involve multiple regions, individuals, or pieces of content.)

While the information is collected, staff will set a date for a conference for the panelists. Once comments are collected, they will be distributed to panelists.



The panelists should discuss the case in a “Panel Conference,” which would give them an opportunity to verbally discuss the arguments and issues before them. Decisions shall be made by majority vote.

When the panelists have come to a conclusion about how the issue should be resolved, one panelist will be selected to draft the opinion. Once the opinion is drafted, the panelist will circulate it to the other panelists, who may offer suggestions or revisions. If any panelist disagrees with the decision and/or its accompanying opinion, she can draft a dissenting or concurring opinion to accompany the final decision. The opinion should be posted publicly within only a few weeks of the panel conference. Staff will assist the panelists to ensure that the opinion does not include any personally identifying information.

The board may sometimes elect to hear cases en banc—that is, to convene a larger, nine-member panel—to review decisions made by three-member panels. En banc panels would also need to make and publish an opinion as soon as is practically feasible.

As discussed in Section III, the board could consider creating a “crisis channel” for urgent matters. The board should consider carefully the benefits and drawbacks of crisis procedures, including tradeoffs between the need for swift resolution and the procedural safeguards proposed here.

B. Initial Disclosures

1. Announcement

The board should publicly announce on its website an issue it has taken up for review. The issue should be described with enough generality and anonymity to ensure the privacy of all individuals and parties involved. The facts of the issue (e.g., content removed or account left up) should be described in general terms such that it does not identify the poster, target, or details of the content. The announcement must also exclude any personally identifiable information.⁴⁷

2. Comments from Individuals and/or Organizations Directly Involved

The board may invite individuals and/or organizations directly involved in an issue to comment after the issue has been announced. The board must solicit comments from the original poster of the content and others affected by the content (e.g., specific targets of harassment, named or seen individuals in a video, affiliated groups, etc.).

Board staff should reach out to all directly affected parties to ensure their voices are heard. Directly affected parties may fill out a standardized form made available to them through Facebook or via email. If directly affected parties cannot be reached, the board will proceed with its processes. If a directly affected party does not want to participate in the comment process, no further action will be taken by the board to include that party.

Once a decision has been made by the board to review a case, no individual or organization has the power to abort the process. As stated in Section III with respect to case selection, proper confidentiality mechanisms should protect the privacy of involved stakeholders.

⁴⁷ See IV.D.3 for more details on transparency.

3. Comments from Other Interested Parties

The board should solicit comments from groups beyond the specifically interested parties to consider varied perspectives on the issue. Other interested parties may include non-governmental organizations, political groups, religious groups, advocacy groups, and the general public. Comments from these groups might include:

- Descriptions on whether similar content should be permitted, forbidden, etc.
- Suggested decision-making processes to be used for the case in accordance with the Community Standards.
- Suggested changes to the Community Standards based on the content and issue.

To ensure that the ultimate decision is well-informed, after an issue has been announced, the board could open a public portal for comments from individuals, groups, or organizations interested in providing comments on the issue.⁴⁸ Of course, an open commenting system does risk inundating the board with information or skewing the board's perspective where comments overwhelmingly come from a single interest group. Indeed, the FCC's recent efforts to repeal net neutrality are instructive; there, the FCC received over 20 million comments, many of them coming from bots.⁴⁹

The board would thus need mechanisms to sift through a large volume of comments. For example, the board will need to employ filters for spam and frivolous, repeated comments. Staff could then review all other comments, extracting patterns and common concerns from the public. The board might also consider ways to create friction and limits on responses (e.g., page limits for comments, requiring a Facebook account or email address to comment) to cut down on spam.

To ensure a thorough and fair review process, it is important that the board staff review (though not necessarily respond to) *all* comments. The board staff would then provide a written synthesis of all public comments to the panel for its deliberation process. While staff would have discretion in drafting the report, it should generally include the following Sections: (a) summary of the comments; (b) descriptions of the key concerns and recommendations brought forth by the public; and (c) patterns and conclusions from the comments. The report should *not* include any suggestions by the board staff on how to deliberate about the case.

C. Deliberation

1. Panels

All cases should be decided initially by a three-member decision panel, made up of current board members. In contrast to the rules that govern many private arbitration organizations, parties should *not* have the opportunity to select one of the panelists or request particular members from an approved list. Instead, the appointment of board members to particular panels should be automated and randomized.

⁴⁸ One model of notice-and-comment in the rulemaking process is that used by U.S. administrative agencies: See Todd Garvey, Congressional Research Service, *A Brief Overview of Rulemaking and Judicial Review*, March 27, 2017, <https://perma.cc/S6VA-W8DL>.

⁴⁹ Glenn Fleishman, "FCC Chair Ajit Pai Admits Millions of Russian and Fake Comments Distorted Net Neutrality Repeal," *Fortune*, December 5, 2018, <https://perma.cc/AG43-BLRC>.

At least one appointed panelist should be from the geographic region from which the dispute arose.⁵⁰

Conversely, in order to ensure a diversity of viewpoints and to prevent a fragmentation of Facebook's policies by regions, we recommend that no more than two of the three panelists come from the geographic region from which the content originated.

Admittedly, there is a danger that the panelists who are from the country or region from which the post originates will exercise undue influence on their colleagues; the panelists who are not from the same nation or region as the dispute may defer to the panelists who do share that cultural background, assuming that their colleagues are better positioned to come to a reasonable decision. These potential sources of panelist bias likely cannot be eliminated without also sacrificing the effectiveness and knowledge of the panel system. The best solution is to identify and warn against these dangers.

Panelists will typically serve on more than one panel at a time. However, they may accept or decline assignments in order to manage their workload. Panelists should also be permitted to decline particular assignments in instances in which they believe they could not give the matter a fair hearing, either because of a personal, financial, or professional conflict of interest.

The Chair of the Board should develop clear standards for when a matter represents a conflict of interest. For example, if there is a dispute between a student and a university administration with which a board member has an academic affiliation, must that member be recused? Similarly, if a board member has donated money to or volunteered for a political party or non-profit organization in the past, would he or she be permitted to hear a matter involving that party or organization? This report does not offer substantive answers to these questions. However, it is imperative that the board develop clear guidelines. The board could, for example, develop a checklist for "conflicts of interest" that board members can consult if concerned about potential conflicts.⁵¹

2. The Record & Argument

The record should include all relevant material. Panelists should enjoy discretion to admit information into the record that they consider necessary to make a reasoned, well-informed decision. In determining what information is necessary, the panelists should weigh the relevance and probative value of the information against projected costs, delay, and general practicality of gaining access to it. While the panelists have broad discretion to shape the record, they should nevertheless follow a common set of guidelines.

The record should presumptively include the following materials:

- The content in question (e.g., the original post)
- All materials submitted by the directly-affected parties (e.g., the appeals from the original poster)
- Comments from outside parties

⁵⁰ In cases in which a controversy spans two or more geographic regions, the Chair of the Board can determine which nation or region has the most significant relationship to the reported content. The Chair might adopt something akin to the "most significant relationship" test applied under U.S. conflict of laws rules. See, e.g., *Itar-Tass Russian News Agency v. Russian Kurier, Inc.*, 153 F.3d 82 (2d Cir. 1998) (applying the test in a cross-border copyright infringement case).

⁵¹ The board might consider models from other dispute resolution systems for "conflict checks." See, e.g., International Bar Association, *IBA Guidelines on Conflicts of Interest in International Arbitration*, 2015, <https://perma.cc/2D9R-DHQM>. The IBA guidelines include "red," "orange," and "green" levels of potential conflicts.

- An operational record from Facebook (e.g., when and why the post was taken down, how much it had been shared, what happened on appeal, etc.)
- A statement from Facebook that includes:
 - Analysis of the content as governed by the Community Standards as currently interpreted
 - Questions that Facebook requests the panel to answer
 - At its option, Facebook’s recommended approach.

The board might also consider establishing a “User Advocate,” who would be independent from Facebook and funded alongside the board. Original posters or account holders for issues under review may choose not to participate in a review. Some users might not provide significant background or justification for why a certain policy should be adopted. Some users might not respond in time. Others simply might not care. In those circumstances, a User Advocate could step in and defend the best case position for the user. The User Advocate would be, in a sense, the representative for all Facebook users. We recommend that Facebook study and implement the best practices of other sectors where consumer advocates have been in place for decades, such as in electrical utility regulation.⁵²

The board should also have access to briefings by relevant subject matter experts. The Decision Panel may ask for briefings on the social, political, and cultural context implicated by a given matter. We recommend that Facebook maintain a roster of subject matter experts who can be called on to provide briefings or consultations on particular matters.

As a general rule, the expert briefings should come from independent scholars, researchers, or consultants unaffiliated with the board or Facebook. However, it may sometimes be appropriate for Facebook to refer the board to its own employees. In particular, Facebook data scientists and operational experts can offer unique insights into the prevalence of content on the platform, the typical behavior of users in relation to that content, and the feasibility of using deranking or other remedies, other than content takedown, across the platform.

In many cases, a briefing from a legal expert knowledgeable in international human rights standards and freedom of expression will be especially important. Keeping in mind that the board is not a judicial body, it is nevertheless essential that its members consult with legal scholars with expertise in international human rights standards, and with particular knowledge in laws related to freedom of expression, access to information, freedom of assembly, and other rights implicated by online content moderation.⁵³

The board can invite additional interest groups and non-governmental organizations to present written arguments to the panelists. Like *amici curiae* in a U.S. federal court, outside interest groups can provide valuable substantive knowledge to the decision-makers and ensure that the panel is considering the best possible arguments.

Given the sensitive nature of much of the content at issue, the record will generally not be public. The board will only share descriptions of the relevant issues, being careful to withhold or redact personally identifiable

52 See Adam R. Fremeth, Guy L. F. Holburn & Pablo T. Spiller, “The Impact of Consumer Advocates on Regulatory Policy in the Electric Utility Sector,” *Public Choice* 161, no. 1-2 (2014), 157-181, <https://perma.cc/N36D-5B9X>.

53 See David Kaye, A/HRC/38/35, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, April 6, 2018, <https://perma.cc/EZA9-P72J>.

information. Pre-screening of outside groups and rigorous non-disclosure agreements may sometimes be needed to safeguard the privacy of directly affected parties as well as the general integrity of the deliberation process. In some exceptional cases, it may not be appropriate to solicit opinions from outside parties at all, despite the general presumption in favor of allowing them to participate.

3. Governing Standards & Principles

The board should articulate the standards and principles it uses in reviewing issues.

Facebook's Community Standards should presumptively govern the board's decision, as should any "Guiding Principles" or Constitution-like document that the board develops (as described in Section I). The Community Standards are the guiding document for Facebook's internal policy team and guide the decisions made by content reviewers. The Community Standards also publicly set expectations and rules for Facebook users. When the board reviews an issue, members should first review the Community Standards to see how the issue might be resolved based on the text, as understood in the light of the context in which it was drafted.

In many cases, however, the Community Standards alone will not suffice. Where the Community Standards are vague, unreasonable, or unfounded, the board must look elsewhere. The board should look first and foremost to its "Guiding Principles" or Constitution (to the extent they exist). Where Community Standards conflict with the board's "Guiding Principles," the board can and should recommend that the Community Standards themselves undergo revision.

D. Decision

1. Opinions

The outcome of this process will be the board's decision, accompanied by an opinion. Generally, the board's decision will apply and interpret the Community Standards in relation to a piece of content, although other recommendations may flow from the board's decision as well (see Section V).

While the board has significant freedom to determine what standards and principles it should use to make its decision, its written opinion should explicitly state what standards and principles it has considered and how they apply to the issue at hand. In its written opinion, the board should cite to any authorities, sources, and experts used to make determinations, explaining their relevance to the decision. This transparency will contribute to the board's legitimacy. And its reasoning may provide Facebook users with guidance about why the board made its determination and how similar posts might be viewed in the future.

The opinions should not be steeped in legal jargon or require expertise to understand. Rather, anyone reading the opinion should be able to understand the central issue reviewed, why the panelists came to the determination they did, and what this decision means for the future of content (or accounts) similar to the one at issue. Although the board may want to describe the issue in the opinion in general terms, it should ensure the opinions do not include any personally identifying information prior to their publication. The names of specific panelists should not be included on the opinion, but rather the opinion should be signed by the board as a whole. After a decision is made, the relevant Community Standard should be publicly annotated with a citation to the board's opinion.

2. En Banc Review

The structure of rotating, three-member panels allows the board to review a greater number of cases than it could with larger panels or by having the full group weigh in on each case. At the same time, the three-member panels may be ill-positioned to recommend significant changes to Facebook's content policies without the input of their colleagues. Moreover, some issues may be especially divisive and require input from more than just three board members. Nevertheless, asking for review by the full board may be unduly cumbersome and would substantially limit the number of cases that the board could consider each year. Therefore, we recommend that the board develop a system of en banc review using a larger (e.g., nine-person) panel, randomly selected from the board membership.

We recommend that if the three-member panel decides it is divided about an interpretive question that reaches beyond the issue at hand, a petition for en banc review may be filed. Upon receipt of a petition for en banc review, a majority of the en banc panel will determine whether they should re-hear a case that led to the original panel's decision.

3. Transparency

The board should view its constituency as social media users rather than Facebook itself. The board should strive to provide as much transparency and access to information publicly as possible. In addition to creating an initial set of rules promoting transparency, the board should develop a process by which it revisits its transparency policies in order to ensure that its approach adapts to the evolving nature of the board's tasks.

The board should have an online presence distinct from Facebook. The website should operate in multiple languages. The website should include basic information about the board, its procedures, and biographies of its members. The board might consider using government or international organization websites as a model for the level of transparency it should provide.⁵⁴ A website would provide significant benefits to the public: it would allow users to understand the process of appealing to the board as well as access to information about prior decisions.

The board's website should also include sections for issues pending review and decisions rendered. The "pending review" section should include descriptions of all issues under consideration (without personally identifiable information). It may also include the option for public comments on the issues. The "decisions rendered" section should include the final written opinion for every case.

The board should exercise caution in describing the content and affected users. Any issue that is reviewed by the board will be magnified on a global scale; it may receive significant press and public attention. Users can often be identified based on a few pieces of information, even when they seek to remain anonymous.⁵⁵ While this might be less concerning for public figures, private individuals have a right to privacy even if they post quasi-publicly on Facebook.

54 See, e.g., Court of Justice of the European Union, <https://perma.cc/JW6E-WU7U>; Conseil Constitutionnel (France), <https://perma.cc/QFY5-YSNM>; International Court of Justice, <https://perma.cc/GKZ8-Q9DJ>; Supreme Court of India, <https://perma.cc/8JUX-VLQW>; Supreme Court of the United States, <https://perma.cc/CPH8-MJPG>.

55 Natasha Singer, "With a Few Bits of Data, Researchers Identify 'Anonymous' People," *New York Times*, January 29, 2015, <https://perma.cc/P65F-63KS>.

The board must be cognizant of users' significant interests in privacy and control of their identity. The online community can, at times, "dox" (or identify) users, harass users, and shame users who would have otherwise wanted to remain anonymous. As such, the board should consider what risks it takes in providing contextual user information when writing an opinion and describing an issue.

In our view, the board does not need to post the piece of content (text, photo, video, or otherwise) as part of its issue announcement or written opinion. Rather, a brief, abstracted description of the content at issue should suffice in most scenarios. If the board needs to provide information that could be traced back to the individuals involved in the posting, or if even an abstract description of the content would reveal their identities (e.g., with especially high-profile content), the board must obtain consent of all individuals. This could pose a significant hurdle for transparency in the decisions, which is why the board might err on the side of providing less factual background and more policy-based reasoning in its decision.

V. Implementation of Board's Decisions

This Section considers how Facebook can ensure that the board is effective in carrying out its oversight mandate, as well as enhance the board's legitimacy in the eyes of users and other stakeholders.

We note at the outset that our concerns about the board's authority and the implementation of board decisions are motivated by a longstanding critique of the effectiveness of judicial review in contexts analogous to Facebook's. In the context of agency mass adjudication, leading scholars have documented that case-by-case judicial review can be ineffective at improving systemic problems of mass decision-making.⁵⁶ A court may provide lengthy reasoned decisions for isolated cases, but such decisions may have little effect on the accuracy and consistency of frontline decision-making due to sheer caseload pressures. In Facebook's context, these challenges strike us as acute: each of some 30,000 content moderators makes decisions once every few seconds, providing little opportunity for deliberation.

We consider six dimensions that affect the implementation of precedential board decisions. First, we consider the scope of the board's decisions, including whether the board can recommend correcting flaws in Facebook's content moderation policies and procedures that are more likely the source of systematic error. Second, we consider what remedies the board can issue beyond reversing a specific decision. Third, we address the extent to which the board's decisions will be binding on Facebook, including whether the board should have the capacity to issue advisory opinions outside the context of an actual content dispute. Fourth, we consider whether the board's decisions will be binding on the board itself—that is, whether decisions will carry precedential weight. Fifth, we address the board's powers to monitor compliance with its decisions. Finally, we discuss how an annual report by the board could serve the board's oversight mandate.

A. Decision Scope

Facebook can constrain the board's decisional purview to various degrees. The Draft Charter proposes the board focus on reviewing moderator decisions. But in reviewing such frontline decisions, the board will develop insights into the Community Standards and other Facebook policies and procedures. In some cases, the board may conclude that Facebook's policies are themselves inconsistent or unfair, and not just their application to specific disputed content. In such cases, the board should not restrict its decision to specific content decisions rather than addressing underlying flaws or inconsistencies. Indeed, it would undercut the board's perceived legitimacy by suggesting Facebook created the board to offload criticism over tricky, one-off decisions rather than to invite meaningful scrutiny.

To achieve its mandate then, the board needs defined authority to assess a discrete subset of Facebook's policies, procedures, and processes. The relevant basket of policies and procedures will likely change as Facebook's content moderation system itself evolves, but should include the Community Standards and, ideally,

⁵⁶ See, e.g., Daniel E. Ho, Cassandra Handan-Nader, David Ames, and David Marcus, "Due Process and Mass Adjudication: Crisis and Reform," *Stanford Law Review* 72 (2019, forthcoming), <https://perma.cc/G8XW-4N7C>; David Hausman, "The Failure of Immigration Appeals," *University of Pennsylvania Law Review* 164 (2016), 1177-1238, <https://perma.cc/EB9U-XLW2>; Jerry L. Mashaw, "How Much of What Quality? A Comment on Conscientious Procedural Design," *Cornell Law Review* 65, no. 5 (1980), 823-835, <https://perma.cc/XX79-VXLD>.

content moderator training and perhaps even algorithmic design. Reversing individual take-down or leave-up decisions as inconsistent with the Community Standards is counterproductive if the Community Standards themselves are flawed or if the procedures for enforcing them are ineffective.

Mark Zuckerberg recently suggested the need for third-party oversight for social media standards.⁵⁷ And the Draft Charter tentatively proposes a wider scope for the board’s review authority in suggesting that “Facebook can incorporate the board’s decisions in the policy development process.”⁵⁸ Accordingly, a number of civil society organizations have called for the board to play “a meaningful role in developing and modifying policies.”⁵⁹ The board’s legitimacy was a top concern:

*Providing the board with such policy-setting authority would also help legitimize the board, and ensure it is not viewed as simply a mechanism for Facebook to shirk responsibility for making challenging content-related decisions.*⁶⁰

Endowing the board with authority to review Facebook policies and practices — including, at minimum, the Community Standards — demonstrates that Facebook takes seriously its need for independent review. As discussed below, giving the board authority to review policies and procedures is not the same as giving it power to mandate changes.

B. Remedies

When the board finds that a piece of content violates the Community Standards, what “remedies” can it issue? Can it order remedies that reach more broadly than the particular content in question?

We recommend that Facebook empower the board to issue a wide range of remedies for content violations, recognizing that content deletion is not always the best solution. Even though only content-related takedowns—and not, for example, News Feed rankings—should be appealable to the board, the board should not be so constrained in its choice of remedy. The board should also be able to suggest demoting content in the News Feed algorithm, providing greater disclosure as to source, hiding part of the content (by blurring faces, for example), or other measures short of taking content down.

The board should also be able to use individual cases to recommend changes to the constantly evolving Community Standards (although, as discussed below, those recommendations would be non-binding). For example, the board could provide specific language to update the Standards or ask Facebook to clarify language in a particular Community Standard.

Finally, the board may suggest improvements to content moderator review processes and training or to algorithmic updates, given that changes to these processes will be necessary for effective implementation of the board’s decisions. For example, the board might recommend content moderator training after every update to the Community Standards and recommend the inclusion of certain training topics. Similarly, for algorithms,

⁵⁷ Mark Zuckerberg, “Mark Zuckerberg: The Internet needs new rules. Let’s start in these four areas,” *Washington Post*, March 30, 2019, <https://perma.cc/Z3JD-JU7W>.

⁵⁸ Facebook, *Draft Charter*, 3.

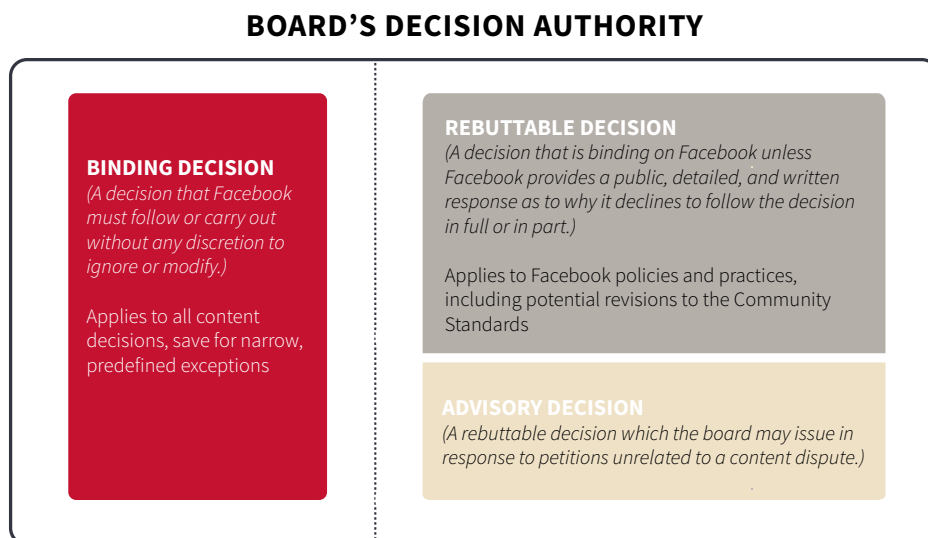
⁵⁹ AfroLeadership et al., *Facebook’s Draft Charter: Joint Statement*, May 2019, <https://perma.cc/BX2S-FUAH>.

⁶⁰ *Ibid.*

the board might suggest specific goals and factors to consider for algorithmic updates, leaving Facebook to coordinate with engineers to update technologies accordingly.

C. Decision Authority on Facebook

We now address whether board decisions should be binding on Facebook with respect to content and the Community Standards.



Our recommendation is three-fold. First, the board's decisions should be binding with respect to content decisions, save for narrow, predefined exceptions. Second, the board's decisions should be rebuttable with respect to Facebook policies and practices, such that Facebook must publicly articulate why it declines to adopt any decision in whole or in part. Third, the board should have authority to provide Facebook with advisory decisions on policy matters, including potential revisions to the Community Standards.

By **binding decision**, we mean one which Facebook must follow or carry out without any discretion to ignore or modify. By contrast, a **rebuttable decision** is binding on Facebook unless the company articulates — in a public, detailed, and written response — why it declines to follow the decision in full or in part. Finally, an **advisory decision** is a type of rebuttable decision which the board may issue in response to petitions unrelated to a content dispute, such as in response to Facebook proposals for revisions to the Community Standards.

Facebook agrees that the board's decisions should be binding as to content.⁶¹ In announcing the idea for an independent review body, Mark Zuckerberg indicated its decisions “would be transparent and binding.”⁶² Similarly, the Draft Charter indicates board decisions will be “binding on the specific content brought for review.”⁶³

61 Facebook, *Draft Charter*, 3.

62 Mark Zuckerberg, “A Blueprint for Content Governance and Enforcement,” Facebook, November 15, 2018, <https://perma.cc/4WBK-Z33P>.

63 Facebook, *Draft Charter*, 3.

We agree that Facebook should be bound by the board’s content decisions, both to ensure the board’s efficacy at correcting errors and to maintain the board’s public legitimacy. Facebook must carefully consider circumstances under which it declines to follow a board decision. The Draft Charter suggests one such circumstance by indicating the board “will not decide cases where reversing Facebook’s decision would violate the law.”⁶⁴ Cases may nonetheless arise where the board makes a decision that might direct Facebook to violate laws in some countries, particularly in countries with restrictive speech regimes. Other unforeseeable cases may arise where Facebook’s other obligations, such as to its shareholders, may contravene the board’s decisions. In these circumstances, Facebook must publicly articulate in writing how implementing the board’s otherwise binding decision contravenes Facebook’s legal obligations or capacities.

We believe binding authority will lend legitimacy to the board and credibility to Facebook. Chris Hughes’ sharpest criticism of the board proposal was that Facebook’s compliance would be voluntary.⁶⁵ By binding itself to follow the board’s decisions on content, Facebook can partially mitigate this critique. In addition, by giving the board binding authority, Facebook signals to stakeholders — including board members themselves — that the oversight board is a worthwhile and meaningful project.

Yet a binding decision alone is not enough if Facebook can easily render that decision moot by revising or eliminating the relevant Community Standard. For this reason, the board should have higher-order, governing principles that cannot readily be revised (see “Governing Principles,” Section I). Moreover, Facebook may need to change its process for revising the Community Standards, perhaps by decreasing the frequency of revisions or by involving the board in the revision process. Indeed, Facebook might seek the board’s input on proposed revisions via an advisory opinion, as discussed above.

That said, there are certain advantages to non-binding decisions. Non-binding international treaties, for instance, can provide parties valuable flexibility by avoiding haggling over specifics.⁶⁶ Similarly, the rise in non-binding mediation as an alternative to civil litigation suggests there is value in having a third party issue a reasoned opinion, even if it does not have binding weight.⁶⁷ Some bodies have a quasi-binding authority that combines the advantages of both. Of particular relevance is the model of the inspector general.⁶⁸ Government agencies generally are not bound to adopt an inspector general’s recommendations, but they are required to respond to such recommendations in writing.⁶⁹ Based on this literature, the board can effectively oversee Facebook’s policies and practices via targeted, non-binding decisions.

We recommend that the board’s decisions regarding Facebook policies and practices be non-binding but require Facebook’s detailed response. This arrangement gives the board the flexibility to propose structural

⁶⁴ *Ibid.*

⁶⁵ Chris Hughes, “It’s Time to Break Up Facebook,” *New York Times*, May 9, 2019, <https://perma.cc/79PW-XVC7>.

⁶⁶ Gregory Shaffer and Mark Pollack, “Hard vs. Soft Law: Alternatives, Complements and Antagonists in International Governance,” University of Minnesota Law School Legal Studies Research Paper Series Research Paper No. 09-23, 713, <https://perma.cc/PE39-F7CU>

⁶⁷ Jeffrey A. Brimer et al., “Using Negotiation and Mediation to Resolve International Franchise Disputes,” *International Journal of Franchising Law* 11, no. 6 (2013), 3-26.

⁶⁸ Glenn Fine, “Seven Principles of Highly Effective Inspectors General,” Center for the Advancement of Public Integrity, May 2017, <https://perma.cc/GD4U-E5SQ>.

⁶⁹ See Kathryn A. Francis, Congressional Research Service, *Statutory Inspectors General in the Federal Government: A Primer*, Jan. 3, 2019, <https://perma.cc/VYX6-PGAN>. For an example at the municipal level, see New York Police Department, Office of the Inspector General, *NYPD’s Official Responses to Reports from the Office of the Inspector General*, <https://perma.cc/HV4R-7B5Y>.

changes and allows Facebook to balance them against fiscal, legal, and other considerations. In turn, Facebook can bolster both its own credibility — and the board’s legitimacy — by adopting the board’s recommendations despite having no binding obligation to do so and by articulating its rationale for declining particular recommendations.

Finally, the board should have authority to issue advisory opinions when Facebook seeks guidance or interpretation outside of a given content dispute. The Draft Charter envisions such a capacity for the board, in that Facebook “may request policy guidance from the board.” There is wide debate over whether adjudicative bodies should issue advisory opinions in the absence of an actual dispute — U.S. federal courts, for instance, cannot, while courts in other countries and certain U.S. states do.⁷⁰ The key advantages of advisory opinions is their proactive nature and flexibility, since they offer answers to broad, complex questions without awaiting claims, at considerable savings in time and cost.⁷¹ Critics note that many disputes are highly context-specific, such that issuing opinions in a factual vacuum is inappropriate. But the evolutionary speed for online content militates in favor of giving Facebook and the board the opportunity to be in direct dialogue about emergent issues. This is particularly true given that Facebook regularly revises the Community Standards, which has the potential to undercut past board decisions.

D. Precedential Value of Board Decisions

In addition to deciding whether the board’s decisions should bind the company, Facebook must decide the extent to which decisions will bind the board itself. Does a given decision bear only on the current content dispute, or does it contribute to a precedential body of “jurisprudence” that the board must draw from in deciding future cases?

Facebook’s Draft Charter provides that the board should consider “consistency with other issued opinions” as a guidepost in reviewing content disputes,⁷² suggesting Facebook envisions some precedential authority. We recommend that Facebook explicitly give the board authority to determine the extent of the precedential authority of its content-related decisions. This retains the board’s flexibility to cabin particular decisions that are incredibly context-specific. In addition, we recommend that the board have the authority to review previous decisions en banc to overturn precedent that has proven incorrect or unworkable.

Precedential decisions are generally thought to promote consistent application of standards, since decision-makers must consider not only the facts and parties before them but also the precedents and potential ramifications of a given decision on future cases.⁷³ Similarly, precedential decisions give parties — such as Facebook users and content moderators — interpretive benchmarks to gauge whether posts violate the

70 Lucas Moench, “State Court Advisory Opinions: Implications for Legislative Power and Prerogatives,” *Boston University Law Review* 97 (2017), 2243-2301, 2246, <https://perma.cc/7SF5-SQRS>.

71 James R. Rogers and Georg Vanberg, “Judicial Advisory Opinions and Legislative Outcomes in Comparative Perspective,” *American Journal of Political Science* 46, no. 2 (April 2002), 379-397, <https://perma.cc/NU69-8BNW>.

72 Facebook, *Draft Charter*, 5.

73 See generally John Hanna, “The Role of Precedent in Judicial Decision,” *Villanova Law Review* 2, no. 3 (April 1957), 367-384, 377-378, <https://perma.cc/B6BJ-R5AV>; Daniel B. Levin, “Fairness and Precedent,” *Yale Law Journal* 110, no. 7 (2001), 1295-1302, <https://perma.cc/XCM8-UCMD>; Luca Anderlini et al, “Legal Efficiency and Consistency,” Georgetown University Working Paper, July 2018, <https://perma.cc/X7SP-CLZ6>.

Community Standards. Precedent also serves the logistical purpose of eliminating the need for the board to hear similar cases multiple times.

Precedential decisions have their drawbacks, however. Notably, precedential decisions might lock the board into bad decisions, until the board is willing to review them en banc.⁷⁴ By contrast, non-precedential decisions have the advantage of simplicity. If the board is only deciding whether a certain post should be taken down or stay up, the panel has no need to consider unforeseen consequences on future cases. Reviewing each case afresh also gives the board considerable flexibility to tailor each ruling to the specific facts of a given appeal.

Our recommendation reconciles the aims of consistency and flexibility. We draw on such models as the U.S. Merit Systems Protection Board, which hears appeals regarding government employee terminations and primarily issues non-precedential decisions but has authority to issue precedential ones, too.⁷⁵ With latitude to choose which of its decisions have precedential authority, the board can marry the procedural fairness of adjudication with the flexibility of an administrative process.⁷⁶ This maximizes the board’s capacity to correct individual content moderation errors — including errors that are so context-specific as to be of minimal precedential value — and also to decrease future content moderation errors by authoritatively interpreting the Community Standards to reduce ambiguities.

E. Monitoring

Effective and legitimate oversight can only be achieved if Facebook actually adheres to the board’s decisions. A major hurdle to compliance with the board’s decisions, and more generally with the Community Standards, is the volume of content on Facebook’s platform and the operational complexity of reviewing this content. An effective monitoring function would keep the board informed about Facebook’s challenges and about the level of compliance with the board’s decisions. We believe that, to be effective, the board will need to monitor three different areas: (1) updates to the Community Standards, (2) content moderator adherence to board decisions, and (3) algorithmic adherence to board decisions.

1. Updates to the Community Standards

As Facebook itself acknowledges, the Community Standards “continuously evolve.”⁷⁷ While these changes are necessary given the dynamic nature of social media content, these revisions risk adversely affecting compliance with the board’s decisions. The revisions might flatly contradict board precedent or, more likely, might result in new content moderation rules that indirectly undermine the board’s existing understanding of the Community Standards. While we considered the possibility that the board would not proactively monitor updates to the Community Standards, that would leave the board in the dark about whether Facebook is adhering to its decisions (unless of course Facebook sought an advisory opinion from the board on a proposed Community Standards revision). We also considered the possibility of requiring Facebook to report on all Community

⁷⁴ Brandon J. Murrill, Congressional Research Service, *The Supreme Court’s Overruling of Constitutional Precedent*, September 24, 2018, 8, <https://perma.cc/TXZK-C4NB>.

⁷⁵ 5 C.F.R. § 1201.117, “Board decisions; procedures for review or reopening,” 1994, <https://perma.cc/S4U7-WA2U>.

⁷⁶ Robert Vaughn, “The Opinions of the Merit Systems Protection Board: A Study in Administrative Adjudication,” *Administrative Law Review* 34, no. 1 (Winter 1982), 25-58, <https://perma.cc/94QX-NLND>.

⁷⁷ Facebook, *Community Standards Enforcement Report*.

Standards revisions, using metrics defined by the board. However, this risks inundating the board with too much information, potentially converting its Community Standards review function into a rubber stamp.

Instead, we recommend that the board have the authority to order periodic or ad hoc reports on revisions to the Community Standards, as the board deems necessary. For example, these reports might discuss the impact of Community Standards revisions on the accuracy and consistency of content reviewer and algorithmic decisions. These reports would thus complement and inform the board's power to recommend revisions, via rebuttable order, to the Community Standards and to issue advisory opinions at Facebook's request.

2. Content moderator adherence to board decisions

To ensure effective implementation of the board's decisions, the board must also stay abreast of how content moderators are implementing their decisions on the ground. Because Facebook relies heavily on human reviewers, who review millions of pieces of content every year, there is a significant risk that the board's decisions will be implemented inconsistently or not at all.

To combat these risks, we recommend that the board be able to order periodic reports related to content moderator review (including the internal appeals process) and training. In doing so, the board may specify which metrics to track, as well as suggest improvements to Facebook's internal enforcement protocols.

3. Oversight of Algorithms

Most of Facebook's content moderation is done through automated filtering rather than human decision-making.⁷⁸ If the board is to have power over content moderation policies, it must have some power to review those algorithms and recommend changes. To be sure, reviewing algorithms and recommending changes to them requires a level of technical expertise beyond that possessed by many who might serve on the board. Nevertheless, to better understand exactly how Facebook is controlling and magnifying the content on its platform, some oversight of the algorithms must be within the purview of the board. Moreover, making recommendations as to the reach of certain content may serve as an intermediate measure between taking down content and leaving it up. Many details as to how algorithms could be reviewed or tweaked will need to be worked out. But oversight of the algorithms must be within the purview of the board.

F. Annual Report

Board members will develop a unique perspective on Facebook's rapidly evolving content moderation system. An annual report would help to distill the board's insights from a wide array of cases and would ensure open communication between Facebook and the board about the overall efficacy of Facebook's content moderation policies.

⁷⁸ For example, while automated filtering algorithms are responsible for take-downs of 96% of all nudity in violation of the Community Standards content, they are responsible for the removal of just 65% of hate speech content. See Cade Metz and Mike Isaac, "Facebook's A.I. Whiz Now Faces the Task of Cleaning It Up. Sometimes That Brings Him to Tears," *New York Times*, May 17, 2019, <https://perma.cc/M5JD-U68E>; Mark Zuckerberg, Facebook post, May 10, 2019, <https://perma.cc/J25V-V8XL>.

The Chair of the Board should shepherd the annual report process. While the content of each annual report should be left to the board’s discretion, reports might include: a summary of the board’s recent decisions, reflections on Facebook’s successes and failures in implementing those decisions, recommendations for improvements to the Community Standards and Facebook’s content moderation policies, and/or lessons for other social media platforms.

Facebook would undoubtedly stand to benefit from the insights of the board. So would the board, whose legitimacy depends in large part on its ability to check Facebook and assure members of the public that their concerns are being addressed. But the benefits of an annual report extend beyond the Facebook platform. As the first real attempt at a social media oversight board, Facebook’s board will face—and hopefully overcome—unexpected challenges, with far-reaching consequences for other social media platforms. It would be unfortunate not to memorialize those lessons, especially as other social media platforms contemplate oversight boards and seek to improve their content moderation policies.

Conclusion

Facebook’s efforts to build a social media oversight board are now well-underway. The company has solicited and reviewed feedback from hundreds of different communities globally, as summarized in its “Global Feedback” report,⁷⁹ and is preparing to launch a board in the near future.

Our findings and recommendations reinforce an emerging consensus about what the board should be. Like other commentators, we prioritize independence, transparency, and diversity for an external oversight board. And, as input across international communities has revealed, it is critical that the board have the power to influence Facebook’s policy development. Yet, just as the Global Feedback report found no “silver bullet” in solving the dilemma of an effective board, this Policy Lab research report reveals the tradeoffs and ambiguities in policy choices, ranging from the size and composition of the board to modes of appeal. Developing an oversight board is no simple task, but doing so is crucial to fulfilling the principles of responsible free expression that allow social media to realize its full potential while avoiding now well-recognized social harm.

Throughout the board development process, Facebook has considered an array of oversight models, seeking one that is effective in guaranteeing due process, the validity and salience of the information considered, procedural fairness, transparency, and efficient disposition of cases. The proposal we have crafted here accommodates these diverse values. Our report highlights the costs and benefits relevant to our proposal and details a plan for implementation.

We applaud Facebook for its willingness to address the problems of content moderation and develop an appellate mechanism for content moderation that has strong institutional legitimacy. As other social media platforms contemplate reforming their content moderation policies, Facebook has the unique opportunity to lead by example. If designed poorly, the board runs the risk of serving as a quasi-judicial rubber stamp on a system plagued by deep structural problems. If implemented well, the board may help to institute genuine reform and systematic improvement, based on transparency and accountability.

⁷⁹ Facebook, *Global Feedback*.

Appendix

A. Example Hate Speech Petition

Below is an example of a structured petition form for the Content Standards Hate Speech category — chosen because this category generates the highest rates of appeal.⁸⁰ Please note that a user will have to log in to their Facebook profile to submit this form for identity and authentication purposes; thus, the form does not explicitly ask for a user name and link to Facebook profile. If a user’s Facebook profile is disabled, they will still be able to submit the petition and their associated Facebook profile will be logged for the case selection process. Ideally an individual without a Facebook profile (i.e., someone victimized by content that does not have a Facebook profile) would also be able to submit a petition for review by the board, but we are not aware of technical mechanisms available to authenticate individuals and prevent spam and impersonation. To prevent spamming, we also recommend that each Facebook account be capped at a specified number of active petitions at any given time. Finally, we recommend the inclusion of a question regarding the privacy expectations of the petitioner because this can serve as a helpful piece of information in the case selection process (for example, if there is an interest in having precedent announced and explained publicly) and because it could inform how much information is included when the board publicizes whether or not a particular case is referred to a Decision Panel.

⁸⁰ Facebook, “An Update on How We Are Doing At Enforcing Our Community Standards,” Facebook Newsroom (blog), May 23, 2019, <https://perma.cc/H6C5-QNU2>.

HATE SPEECH VIOLATION BOARD PETITION

If the piece of content that you would like the board to consider violated (or is believed to be a violation of) the [Hate Speech provisions of the Facebook Community Standards](#), please fill out this form.

BACKGROUND INFORMATION: First we need to collect some basic information about your petition to the board.

1. Link to the content that was removed or the content that you believe should be removed from the platform: _____
2. Date that content was posted: _____
3. Date that content was removed (or first flagged): _____
4. Was your Facebook account disabled as a result of this piece of content: [Yes | No]
5. Did you post this content? [Yes | No]
 - a. If so, please explain why you posted this content. Provide some background on who you are.
6. Is your image or name included in the content you are petitioning? [Yes | No]
 - a. If so, please describe your involvement in the creation of the content.
7. Are you victimized by the content? [Yes | No]
 - a. If so, please describe how you are victimized by the content.
8. Did Facebook interpret the [Hate Speech provisions of the Community Standards](#) correctly when they removed this content (or when they decided to allow the content to remain on the platform)?
 - a. Yes, but the Content Standards should be changed because this decision is wrong.
 - b. No, Facebook did not correctly apply the Community Standards to this piece of content.
 - c. No, my issue does not involve an interpretation of the Community Standards.
9. [OPTIONAL] If you believe the Hate Speech Community Standards should be changed, please include an excerpt of the relevant text of the Community Standards and explain how it should be modified and why: _____

10. Do you want to keep this case to be kept confidential? [Yes | No]

CONTEXTUAL INFORMATION: Help us better understand why you think an incorrect decision was made. The more information you can provide in response to the below questions, the better we will be able to assess your petition and its eligibility to be heard by the board.

11. Who is being attacked or targeted by this piece of content? Is it a specific individual?
12. Why is this individual or group of individuals being attacked or targeted? On what account are they being targeted? Are they being attacked or targeted because they possess a particular characteristic (i.e., religion, race, ethnicity, religious affiliation, sexual orientation, gender, a disability, etc.)?
13. How does this content attack or target the individual or group of individuals? Please indicate all that apply and describe what specifically the content does.
 - a. Violent speech
 - b. Dehumanizing speech
 - c. Statements of inferiority
 - d. Expressions of disgust
 - e. Expressions of contempt
 - f. Call for exclusion
 - g. Use of insulting slurs
14. Does this content serve an educational purpose or does it raise awareness about a particular issue? If so, please elaborate.
15. Does your request relate to pieces of content in addition to the one you specified above? If so, describe the scale of this issue and the other pieces of content this case implicates.

Is there any additional information you would like us to know to process your petition?

[Open Text Field]