FACEBOOK'S OVERSIGHT BOARD: WHY FACEBOOK MIGHT SEEK TO 'BIND ITSELF TO THE MAST'

Jessie M Kelly

August 2020

ABSTRACT

Content moderation is the core business of Facebook. This paper seeks to explore the extent to which the new Facebook Oversight Board will influence Facebook's content moderation practices and why it was created in the first place. It begins exploring the buy in from users to the content available on the platform, without which Facebook ceases to have a business model. As such, the extent to which Facebook is seen to legitimately respond to criticisms of its content moderation enables it to retain its position, legitimacy and autonomy. The Facebook Oversight Board or 'FOB' is one exercise aimed at retaining control over content moderation, and while the Board will see only a small subset of cases comprised of those already removed from the platform it will play a broader role in influencing the perception of Facebook's content moderation practices, and subsequently in shaping those practices and associated international legislation.

1. INTRODUCTION

The varied and numerous content moderation pressures under which the executive of Facebook were operating prior to the establishment of the FOB are well documented by both scholars (see, for example, Klonick, 2018, 2020 and Douek, 2019) and directly by Facebook (Bickert, 2019).

Each day Facebook must attempt to address about 275,000 content moderation requests from over 2.5 billion pieces of content shared.¹ The volume of requests and associated public engagement activities represent significant operational requirements for the business and have necessitated a significant increase in investment in artificial intelligence (AI) technology and research. This investment is easily justified as Facebook's ability to convincingly respond to calls for better content moderation in a way which both maintains user trust and

¹ In the first quarter of 2019, Facebook received nearly 25 million requests for appeal. Douek, Evelyn. "Facebook's Oversight Board: Move Fast with Stable Infrastructure and Humility." NCJL & Tech. 21 (2019): 1.

engagement and is enticing to new users is central to continued business growth.²

Accordingly, content moderation has increasingly become core business for Facebook over the last 10 years. Some of the newest products on the platform reflect this.³ Various described as 'the commodity that platforms offer users'⁴, an 'impossible and indispensable part of Facebook's business'⁵ and the 'business reasons'⁶ for the push for external self-regulation which resulted in the FOB, content moderation has become an ever present and time consuming requirement for Facebook.

Or, as Mark Zuckerberg put it to US Congress on 29 July 2020 in testimony to the anti-trust commission:

Senator, having hate speech on our platform does not help our business, it harms our business.

A key issue in Facebook's evolving content moderation practice has been maintaining public trust and acceptance of their methodology. To retain users on its platform and attract new ones Facebook seeks to solve content moderation issues to a broadly accepted degree. They must respond convincingly to dissatisfactions and criticisms of inconsistent and opaque processes in their content moderation practices.⁷ Most pointedly this requires addressing safety and security concerns but also developing a further nuanced position on what

² Klonick, Kate. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression." Yale LJ 129 (2019): 2418.

'Facebook's ability to draw and keep users on its site in order to view advertising. Users' decisions to use Facebook depend on their trust of the platform. Facebook's creation of the Oversight Board is an investment in building user trust, which is a long-term strategy for continued economic growth.'

³ Namely, Facebook's effort to register 4 million voters for the US 2020 national election via push notifications and a 'Voter Information Centre' and the 'COVID-19 Information Centre' incl. community building and support efforts.

⁴ Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the hidden decisions that shape social media*, 2018, p. 5 cited in Douek, 2019.

⁵ Douek, 2019

⁶ Monika Bickert in Douek 2019, 'But as Facebook's Head of Policy Management has written: "[s]imply put, there are business reasons that a big social media company must pay attention to what the world thinks of its speech rules."'

⁷ Klonick, 2019, 'The rise of public awareness and demands for legality around platform content-moderation governance empowered civil society, academics, and journalists'; 'Although details about Facebook's content-moderation rules and how they were enforced may have been hidden from those outside the platform for years, that does not mean they were not free from external influence. To the contrary, many of the policy and enforcement decisions made by architects inside Facebook came about in reaction to external pressures from civil society, governments, the media, or users. Of these, negative media coverage had arguably the most powerful impact.'

constitutes acceptable speech on Facebook.⁸ History has shown that specific interest groups can have an arguably outsized impact on content moderation, whereby an issue for one group can require a whole of platform response.

Similarly, while the FOB will deal with only a small subset of content moderation issues i.e. taken only from those instances where content has been removed from Facebook and the normal appeals channels have been exhausted, this direct caseload is proxy for their role in a larger conversation and we should expect to see them treat their moderation as such.

Content moderation has become core business and Facebook is responding on multiple fronts; by increasing cooperation with international governments on threats to national security, investment in operational capacity via AI, and a more visible role in international relations. A key part of its strategy is self-regulation through the creation of the FOB which aims to increase the amount of legitimacy that Facebook has to play with.

Describing the conditions leading to the emergence of the FOB Evelyn Douek considers that 'Comparative scholarship shows that "[n]ew constitutional forms emerge only under extraordinary historical conditions, moments when pre-existing political, economic, and social structures have been weakened [or] delegitimized.'" Content moderation practices prior to the FOB received unprecedented public scrutiny and suffered a crisis of legitimacy, as such representing such conditions for online platforms.'⁹

The creation of the FOB is primarily an exercise in maintaining and enhancing the legitimacy of Facebook's content moderation practices/business.

Adopting the court like structure was suggested by Harvard Law Professor Noel Feldman – originally termed 'A Supreme Court for Facebook' – the FOB sources its legitimacy from accepted and transparent legal methods and widely held values, such as authenticity, safety, privacy and dignity as it seeks to strike a balance between effective content moderation and the protection of online free speech.¹⁰ Moreover, the FOB plays a role in creating input for a feedback loop which will be used to train the artificial intelligence (AI) that moderates most of Facebook's content.

⁸ E.g. this has meant the modification of policies and procedure as well as changes to back end systems, including training AI to better recognise art or cultural items such as the 'Terror of War' photo.

⁹ Alec Stone Sweet, *Governing with Judges Constitutional Politics in Europe*, 2000 in Douek, 2019, 38.

¹⁰ Zuckerberg, Mark, "Facebook's commitment to the Oversight Board" Facebook Newsroom September 2019, about.fb.com/wp-content/uploads/2019/09/letter-from-mark-zuckerberg-on-oversight-board-charter.pdf

By creating a self-regulatory external body Facebook also seeks to stave off or guide developing regulation, outsource responsibility and retain market share.

Despite its importance, the FOB is only one exercise in retaining the platform's position, legitimacy, and autonomy. Increasingly Facebook is required to cooperate with multiple national governments on issues of national security and data privacy.

This paper aims to understand why Facebook would chose to self-regulate in this way and the motivations which led Facebook to form the FOB.

2. LEGITIMACY VIA PUBLIC REASONING

The FOB was conceived to alleviate the burden associated with content moderation issues by increasing the amount of legitimacy Facebook was granted to deal with them via a process of public reasoning.

Public reasoning has its roots in Kantian philosophy but is most relevant here as defined by John Rawls (1993) - 'public reason requires citizens to be able to justify their political decisions to one another using publicly available values and standards.'¹¹

The use of political power in a liberal society will be legitimate if it is employed in accordance with an interpretation of the fundamental ideas implicit in a society's public political culture (a political conception of justice, i.e. publicly available values), and which must be realised in certain kinds of institutions.¹²

Citizens engaged in certain political activities have a duty of civility to be able to justify their decisions on fundamental political issues by reference only to public values and public standards.

This is Rawls's doctrine of public reason by which legitimacy is retained and the law may permissibly be enforced.

Content moderation is the 'industry term for a [social media] platform's review of user-generated content posted on its site and the corresponding decision to keep it up or take it down'.¹³

The perceived (complete) lack of legitimacy in Facebook's management of content moderation required an appropriately large response housed in a certain type of institution.

3. THE PROPOSED RESPONSE

¹¹ Wenar, Leif, "John Rawls", The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.), forthcoming URL = <<https://plato.stanford.edu/archives/sum2021/entries/rawls/>>.

¹² Ibid.

¹³ Klonick, 2019

Scholarship on the topic as well as the Facebook 'Global Consultation Report'¹⁴ credit Harvard Professor Noel Feldman with the original concept for an external oversight board.¹⁵

In 2018, Feldman observed asymmetry between specific advocacy groups –in calls for content moderation to which Facebook was required to respond – and general support for free speech online.¹⁶ He advocated that Facebook employ a private court structure. As he saw it government could not regulate in this areas as their responsibilities were to protect free speech for Facebook. And so Facebook should create its own 'Supreme Court'.

As outlined above, Rawls's doctrine of public reason allows for the creation and maintenance of legitimacy in order that the law may permissibly be enforced, and public reasoning is housed in particular institutions. Thus an external court lends the necessary legitimacy to content moderation practices via public reasoning – creating accepted governance of content moderation.

This concept is central to Feldman's design:

*The key to making this private-courts approach work is recognizing that there is no magic-bullet solution to balancing competing values of free expression, safety, and equality. The advantage enjoyed by real-life constitutional courts is that they openly address difficult cases, and so derive credit and legitimacy from being principled. They make mistakes, and correct them. Their rules evolve with changing technology and ideas. And instead of blaming them for this, we mostly validate their efforts.*¹⁷

By announcing the 'reasoning behind their decisions on a case-by-case basis... [facebook's] private "courts" would be borrowing some of the legitimacy of public courts'¹⁸

As such, a desire to establish this legitimacy explains Facebook's creation of self-imposed external regulation.

¹⁴ Darmé, Zoe et al., Global Feedback & Input on the Facebook Oversight Board for Content Decisions, FACEBOOK 41, 27 June 2019, fbnewsroom.us.files.wordpress.com/2019/06. Hereafter 'Consultation Report'. The summary report covering the consultation phase which ran a series of workshops to inform the development of the FOB. The two original papers from Feldman are included as Appendices in the Consultation Report but are otherwise unpublished.

¹⁵ Ibid. - 'the Oversight Board has continued to evolve from the Constitutional Court conceptualized in Feldman's original paper'.

¹⁶ Feldman January 2018 and March 2018 in Consultation Report 'Those seeking to bar speech are issue focused and concentrated. Those seeking to protect speech tend to be diffuse and not issue-focused.'

¹⁷ See Consultation Report Appendix: Noah Feldman, A Supreme Court for Facebook (January 1, 2018). (NB: See Klonick, 2019, 2449, Feldman is reported to have written this after discussion with Facebook senior executive and passed it to Sheryl Sandberg.)

¹⁸ Ibid.

Evelyn Douek summarised this as 'it is the process of public reasoning itself which is the *raison d'être* of the FOB'.¹⁹

AI does not give public reasoning for its decisions, which...is an essential component of a legitimate content moderation system.

*Public reasoning, then, is about facilitating [global] dialogue and legitimizing Facebook's exercise of power over its users.*²⁰

Although there were a range of other advocates for an 'external independent appellate body and a team was building inside Facebook from January 2018'²¹ Facebook credits Feldman's concept as continuing to drive operations stating in 2019 'the primary objectives also remain the same [as in Feldman's original conception]: increased transparency, legitimacy of decision making, and independent judgment.'

Along with Feldman, many of the other experts consulted in the earliest days of the development of the FOB had a formal legal background and/or were resident in America where courts are the primary institutions by which democratic principles are upheld.²²

It appears that for Feldman's protection of freedom of expression was paramount.²³

Fortunately, there is a solution... [to free speech taking a back seat to content based restriction] The principle of free

¹⁹ Douek appears to draw directly on Rawls – i.e. that the FOB is important for the process of public reasoning necessary for persons in a pluralistic community to come to accept the rules that govern them, even if they disagree with the substance of those rules.

²⁰ Ibid.

²¹ Klonick, 2019, p. 2449

²² Kate Klonick spent time in residence with the Facebook team spun up to deliver the FOB. Interviews with the team and others in Facebook's network located outlined to Klonick how the FOB was developed over multiple casual conversations, dinner parties and many informal meetings of American based academics, business leaders and others working in or around Facebook. The effect of these conversations is likely to have had significant impact on the eventual decisions made by executive and escapes explicit record keeping or citation in official documentation outlining the development of the FOB. Klonick, 2019, p. 2438

²³ Facebooks FOB process are deeply embedded in an American history strongly protective of original liberties including protection of free speech. Feldman and others cite this as one of the primary motivators for the Global Consultation process. Further, Klonick (2019) reflects 'Facebook's projected rules and values are not neutral. For example, Facebook has a very vocal predisposition to and commitment to freedom of expression. This is also reflected in its Values, which prioritize [sic] voice over concerns of safety, privacy, dignity, and authenticity.' However, she adds 'Facebook's removal of hate speech, harassment, and bullying, all of which would be permissive in a First Amendment framework, mark concessions to limiting unfettered expression.' Indicate the direction of increasingly interventionist content moderation practices.

expression can be adopted officially and formally as a constitutional rule, the way it is in all democracies.

Following this statement in his January 2018 white paper, he delivered a second, titled '*Facebook Supreme Court: A Governance Solution*' in March 2018. In this second paper, Feldman describes the desired balance that the FOB should strike between 'the values of free expression and association' with 'public safety, equal dignity, and community' through 'principled, reasoned decision making'.

To achieve this he proposes a court comprised of legally skilled paid experts to sit above the existing Facebook content review team. Feldman's model proposed two tiers of independent decision making (carried out by up to 244 independent contractors) with lower independent courts sitting under the Tier 5 Facebook 'Supreme Court'.

In Tier 5 he suggested a range of mechanisms intended to increase the distance between Facebook and the court, the transparency of operations and decision making processes and the legitimacy derived from those decisions. In outlining his reasoning he suggests:

The Court's decisions will be read and analyzed globally. They should cite relevant comparative precedent. Dissent is appropriate and will enhance the quality of decisions as well as public legitimacy.

Beyond that he proposes that judges should be engaged through independent publicly transparent contracts. For Feldman the judicial model is intended to *capture the legitimacy benefits of decisional independence*.

Scholarship on the FOB is, understandably, only just emerging. Evelyn Douek, Marta Maroni and Kate Klonick have been analysing the development of the FOB and potential future implications in real time and all cite the need for Facebook to enhance the legitimacy of their content moderation as, at least in part, the reason for the creation of the FOB.

Douek (2019) cites authoritarian regimes and translates the reasons that these regimes often use courts into the context of the FOB.²⁴ Namely, (1) bestowing content-moderation decisions with an aura of legitimacy, aiding user-relations; (2) staving off or guiding more extensive governmental regulation; (3) outsourcing controversial decisions away from the company; and (4) facilitating better enforcement of existing standards. Much of the perceived benefit is related to reduced cost, including avoidance of more extreme measures of conflict resolution.

Maroni cites Gunther Teubner's Societal Constitutionalism to explain the legitimising and 'self-constitutionalizing tendency' of the FOB whereby a private organization develops its own set or understanding

²⁴ Katsh, M. Ethan, and Orna Rabinovich-Einy. Digital justice: technology and the internet of disputes. Oxford University Press, 2017, 38.

of fundamental rights, which coupled with stronger administrative procedures, reinforce its position, legitimacy and autonomy.²⁵ She concludes that, to the extent that the FOB delivers on Facebook's position, the questionable autonomy of the FOB may ultimately undermine the legitimacy of its theoretical independent structure.

Klonick points in particular to recent erosion of user trust and cites public backlash over Facebook's complete and opaque control over content moderation as key precursors to the creation of the FOB. She posits that the structure that has been created is comparative to a court of limited jurisdiction i.e. administrative courts like the Social Security Administrative Court or the Foreign Intelligence Surveillance Court.

She further highlights a potential increase in legal conflicts as an additional reason why Facebook may seek to self-regulate in content moderation.²⁶

Noting that formal regulation of social media platforms has not kept pace with the development of those platforms and is largely insufficient in providing meaningful guidance on complex issues, it remains the responsibility of Facebook to create legitimacy for its content moderation practices.

4. THE BROADER CONTEXT –LESS CEO BITCH, MORE JAMES GORDON BENNETT²⁷

Considering the bounded reasons why the FOB was established as an external independent body is essential yet the broader context in which the FOB plays a role and which made its establishment more likely also merits consideration. While the primary motivation for the development of the FOB was an investment in legitimacy, the FOB can and will play a role alongside developing conversations on broader discussions of the role of social media in the context of national sovereignty and security.

This is the secondary motivation for the establishment of the FOB proposed by Feldman in his original white paper.

*A further benefit of a quasi-legal approach is that it would give the platforms much greater authority in resisting the demands of countries who want them to restrict speech in ways they would prefer not to respect. Right now, the platforms must engage in a complex, unpublicized calculus about how and when they should obey local rules...*²⁸

²⁵ Maroni, Marta. "Some reflections on the announced Facebook Oversight Board." (2019).

²⁶ Klonick, 2019, 2441, 2427 & 2476

²⁷ Matthew Pressman has compared Zuckerberg to historical newspaper figure James Gordon Bennet, in Kolitz, Daniel, "Which Historical Figure Was The Mark Zuckerberg Of Their Time? " Gizmodo 11 May 2020, www.gizmodo.com.au/2020/05/which-historical-figure-was-the-mark-zuckerberg-of-their-time.

²⁸ Feldman, 2018 in Darmé, 2019.

If Douek's lists her reasons in order of impact then they track this formula as well – firstly, bestowing content-moderation decisions with an aura of legitimacy, aiding user-relations, then secondly, staving off or guiding more extensive governmental regulation.

Prior to and surrounding the development of the FOB, Facebook has been required to respond to the platform's involvement with the Cambridge Analytica scandal, Christchurch mosque shootings, fake or troll pages, especially those originating in Russia, repeated requests for increased transparency for political ads and control over third-party use of the data, a U.K. Parliamentary Report calling Facebook 'digital gangsters' and an American anti-trust investigation into big tech.

Mark Zuckerberg as CEO is increasingly required to adopt a global leadership role in conversations of what constitutes good online governance within modern democracies. Most recently Zuckerberg was seen to appeal to the European Union to embed democratic principles in regulation via the European Commission Digital Service Act. Rather than staving off regulation, Facebook seeks to support the creation of regulation which aligns with its current Community Standards.

*The best antidote to [less democratic models for regulating technology] is having a clear regulatory framework that comes out of Western democratic countries that becomes a standard around the world...*²⁹

As of July 2020, Facebook works with over 50 governments on matters relating to safety and security online. Concerns over threats to user data and malicious use of the platform including interference by foreign governments in elections require Facebook both to participate in public conversations about the role of social media platforms in upholding democracy and to work directly with governments to identify and address these in real time.³⁰

In this context both the case by case advice of the experts sitting on the FOB and the publicly reasoned values in themselves are functionally important for the reputation of Facebook. We should expect to see these values captured, codified and employed in a feedback loop to train content moderating AI as well as to act as 'talking points' for facebook executive.

²⁹ Centre on Regulation in Europe webinar, May 18 2020 cited in Politico, "Facebook CEO warns against spread of Chinese-style internet regulations", South China Morning Post, 19 May, 2020, www.scmp.com/tech/big-tech/article/3084962/facebook-ceo-warns-against-spread-chinese-style-internet-regulations.

³⁰ Mark Zuckerberg testimony to U.S. House. Committee on the Judiciary, Subcommittee on Antitrust, Commercial, and Administrative Law, Online Platforms and Market Power, Part 6: Examining the Dominance of Amazon, Apple, Facebook, and Google, 2141 RHOB, 29 July 2020, judiciary.house.gov/calendar/eventsingle.aspx?EventID=3113.

5. CONTENT MODERATION AS CORE PRODUCT DEVELOPMENT

Facebook faces a formidable challenge: it needs to preserve private power over its policy and product, while ensuring long-term engagement by meeting user expectations of accountability.³¹

Content moderation is core business for Facebook. Unable to remain neutral Facebook increasingly plays a role in active moderation of the content on the platform. This changed role is most evident in the creation of new 'products' in 2020/21.

Facebook's COVID Information Centre, for example, curates users information by pushing information from recognised health services above others. Further, their efforts to register 4 million new US voters for the 2020 election is explicitly non partisan but supports a set of values and activities, namely voting. A developing consciousness within the company of this new role is evident in statements from Chief Product Officer Chris Cox who returned to Facebook after leaving a year earlier.

On June 11, 2020 Chris Cox posted internally then externally about his return to Facebook.

In 2019 I refocused my time, spinning up climate change initiatives, building progressive political infrastructure for this election year, playing with my reggae band, and reconnecting with my family and kiddos. Then 2020 refocused us all, on a public health crisis, an economic crisis, and now a reckoning of racial injustice. The world is unsettled, divided. People are struggling when things were already hard.

I've been following Facebook and I've been encouraged by progress on so many of the big issues facing us.

Our most important decisions and products are ahead of us.

Facebook considers itself an arbiter of public discussion, staff are acutely (perhaps delusional) aware of the impact their policy decisions have on the shape of our future society. The broad role outlined above indicates the mutualism the public has with the future shaping role Facebook has taken and the importance of public acceptance and approval of the degree to which Facebook moderates content on its platform.

6. CONCLUSIONS

In the context of increasingly critical public pressure on Facebook to better address content moderations concerns, Facebook executives were searching for answers that would allow them to address questions more convincingly regarding content moderation. The FOB is one intervention aimed at solving that problem. Primarily, via increasing

³¹ Klonick, 2019

the legitimacy with which Facebook is seen to address these issues through public reasoning facilitated by an external court-like structure.

The specific instrument that is the FOB was born of an American legal liberal perspective initially valuing the protection of 'free speech' above all else – namely in the form of Noel Feldman's white papers. However, the FOB is now couched in a broader context which includes the extent to which the core product of Facebook is under review and an increasingly political role for Mark Zuckerberg as CEO.

This increasingly political role of Facebook foreshadows the future of the FOB. With the FOB one key tool in influencing regulation. Facebook cooperates with more than fifty governments who seek to allay national security concerns through increased legislation. Ultimately, we should expect outputs from the FOB to be used as tested and 'publicly reasoned' values and to influence globally accepted data governance standards and to train the AI which bears the brunt of the content moderation work.

ACKNOWLEDGEMENTS

Special thanks go to Markus Anderljung and Hamish Hobbs at the Future of Humanity Institute, University of Oxford.

BIBLIOGRAPHY

Douek, Evelyn. "Facebook's Oversight Board: Move Fast with Stable Infrastructure and Humility." *NCJL & Tech.* 21 (2019): 1.

Darmé, Zoe et al., Global Feedback & Input on the Facebook Oversight Board for Content Decisions, FACEBOOK 41, 27 June 2019, fbnewsroomus.files.wordpress.com/2019/06./oversight-board-consultation-report-2.pdf.

Gillespie, Tarleton. "Content moderation, AI, and the question of scale." *Big Data & Society* 7.2 (2020): 2053951720943234.

Katsh, M. Ethan, and Orna Rabinovich-Einy. *Digital justice: technology and the internet of disputes*. Oxford University Press, 2017.

Klonick, Kate. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression." *Yale LJ* 129 (2019): 2418.

Kolitz, Daniel, "Which Historical Figure Was The Mark Zuckerberg Of Their Time? " *Gizmodo*, 11 May 2020, www.gizmodo.com.au/2020/05/which-historical-figure-was-the-mark-zuckerberg-of-their-time.

Maroni, Marta. "Some reflections on the announced Facebook Oversight Board." (2019).

Politico, "Facebook CEO warns against spread of Chinese-style internet regulations", *South China Morning Post*, 19 May, 2020, www.scmp.com/tech/big-tech/article/3084962/facebook-ceo-warns-against-spread-chinese-style-internet-regulations.

Sweet, Alec Stone. *Governing with judges: constitutional politics in Europe*. OUP Oxford, 2000.

U.S. House. Committee on the Judiciary, Subcommittee on Antitrust, Commercial, and Administrative Law, *Online Platforms and Market Power*, Part 6: Examining the Dominance of Amazon, Apple, Facebook, and Google, 2141 RHOB, 29 July 2020, judiciary.house.gov/calendar/eventsingle.aspx?EventID=3113.

Wenar, Leif, "John Rawls", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2021/entries/rawls/>.

Zuckerberg, Mark, "Facebook's commitment to the Oversight Board" Facebook Newsroom September 2019, about.fb.com/wp-content/uploads/2019/09/letter-from-mark-zuckerberg-on-oversight-board-charter.pdf.