



Considering Facebook Oversight Board: turning on expectations

May 2019

Facultad de Derecho

Centro de Estudios en Libertad
de Expresión y Acceso a la Información

UP
**Universidad
de Palermo**

Considering Facebook Oversight Board: turning on expectations

The International Criminal Court is popularly referred to as the “world Court”. With Facebook’s announcement to create an appeals, independent body to review the application and enforcement of its terms of service over its 2.7 billion users all around the globe, the term “World Court” probably gained at least one more, new-found meaning.

Since the creation of this body launched there have been advocates for and against it. The common lines among both sides is that 1) this is a self-regulation initiative and cannot replace a judicial stance of control; 2) as it stands it only addresses some of the issues that users have with content moderation; 3) there are more questions than answers regarding nature, scope, design, expectations, etc.; 4) the creation of this Board, although an interesting proposal, should not divert the company’s attention from due process for all content restrictions and curation, and for transparency in rule and decision-making as well as implementation of standards, whether legally imposed or self regulated.

Background and implications to the Oversight Board

After a couple of years of scandal after scandal and a growing set of rules, Facebook is seeking legitimacy in governing what the US Supreme Court has ruled to be “the new public square”. The rules (<http://www.facebook.com/communitystandards>) have gained in complexity and ambiguity over the years and with every addendum (internal guidelines are being changed every week or so, and an overall 2000 changes are included per year approx. in total, according to information provided by FB). Content “flagged” by users to be against those rules raises to a meaningful daily amount.

Per user and government pressures of all sorts, the company has moved content moderation from a responsive to a proactive mode, having automated most of its detection practices (yet unclear how much of its removal practices) for unwanted content and facing extremely complex debates over ethics, corporate social responsibility, liability, damages, and free speech, among others.

The decision to create an independent board or Oversight Board for Facebook, with an open process and consultations is maybe the most dramatic decision that we have seen from FB in the last few years, at least regarding its content policy. And a first step among internet companies to bring in external actors to their decision making processes. After having shifted the rhetoric from a free-speech oriented discourse to a safety discourse over the last few years and having been on a defensive and ever expanding curating role for increasingly complex typologies of content, Zuckerberg’s Blue Print for Content Governance and Enforcement (Nov 2018) seems to mark a new departing point for the company’s approach to content regulation.

For one, Zuckerberg’s note appeals directly to the need for improved legitimacy over governance and deci-

sion-making; Second, there is a commitment to create an external independent review board, whose decisions are binding and public. Every other detail, still unknown. Third, it expressly invites States to define what they expect from a content moderation regime and details at least two concrete initiatives that FB has already engaged with in Europe: 1) the agreement signed with President Macron to work on a new content regulation; 2) their workings towards a new European framework for content moderation and regulation within the next two years. Although FB had anticipated their change in view *vis a vis* legislation during hearings before the US Congress in 2018, this is the first communication that FB sends to its own community officially welcoming and even compelling States to regulate. As Zuckerberg states: “At the end of the day, services must respect local content laws, and I think everyone would benefit from greater clarity on how local governments expect content moderation to work in their countries.”

The recognition that “A full system requires addressing both governance and enforcement” is without a doubt a positive step forward in content moderation debates, particularly those related to transparency. Civil society and academics around the globe had voiced their concerns as to the lack of transparency and the levels of discretion that internet companies enjoyed and have been working with them through the years to improve transparency first over the rules and later onto processes.¹ With the new approach, FB may be able to take procedural and enforcement transparency further while making decisions binding and public.

But maybe what is even more interesting is the policy change *vis a vis* content governance on the platform and what seems to be a new found willingness to “share” the responsibility (or blame) over rule-making. “As I’ve thought about these content issues, I’ve increasingly come to believe that Facebook should not make so many important decisions about free expression and safety on our own” wrote Zuckerberg in November. Human Rights activists have been saying so for years. What changed? And how much did it change?

The creation of a board to serve as an audit to this platform’s decisions on content moderation could strengthen the exercise of free speech online as much as it could hurt it. If the Board is understood as an internal process, intended to serve the company in dealing with complex issues of free speech and its balance with other rights, unify criteria, and help the system converse better with international human rights standards, the results could be more positive for the entire ecosystem. If understood as a replacement system for already weak due process mechanisms within the platform’s decision making structure, and as a body that would create and interpret privately legislated law alone (contract law), it will most probably damage the ecosystem as well as the company. It would also fail the purpose for which it was created. If legitimacy is what FB wants and needs, it can only be built from the dialogue between the private self-regulation norms and existing international human rights standards.

Whether this Board can accomplish legitimacy or not in many ways depends on how the body is structured, the goals that are set for it, the requirements that candidates should meet to be a part of this body, the nature of its decisions (if it is an appeals body in fact), and the transparency and publicity of its decisions and reasoning.

¹ RDR, GNI, to name a few.

Legitimacy in adjudicative bodies:

The concept of legitimacy may be approached from a sociological and a normative dimension and the distinction could be useful to illustrate this point. On the one hand, the more positive the public's attitude towards an institution's right to govern, the greater its popular legitimacy². However, this legitimacy is fragile. Particularly for FB Board, which will not have a reservoir of legitimacy accumulated over a long history to draw upon. On the other hand, legitimacy can also have a normative meaning, referring to whether the claim for authority is well founded³. Given the global challenge that FB faces, the legitimacy of FB Oversight Board must be strengthened on both fronts, but even more so in the latter. Building strong normative legitimacy could provide a standard for judging the Board and deciding if it deserves support. Also, normative legitimacy can influence sociological legitimacy, or perceptions of justified authority, and thereby, the extent to which it will undergird or undercut the work of the FB Board.

There is rich literature on what elements contribute to the legitimacy of adjudicative bodies. There are even concrete writings on the legitimacy of international adjudicative bodies. What determines their legitimacy? Scholars and practitioners have identified three key elements to the legitimacy of international adjudicative mechanisms: 1) fair and unbiased decisions; 2) an interpretation and application of rules consistent with its scope and purpose; and 3) that the body be transparent, independent and infused with democratic norms⁴.

Fair and unbiased decisions need to be at the heart of any adjudicative system, not only international ones, in order for them to be legitimate. "Unbiased" has traditionally been related to independence and hence has focused on nomination and selection processes, quality and soundness of reports, decisions and recommendations, the public discourse and the writings of the tribunal members. All these elements have concrete and relevant definitions dispersed among a vast international and comparative jurisprudence that should not be ignored.

Fairness, on the other hand, according to Prof N. Grossman, "need not entail an equal ratio of rulings in favor or against any given party –in fact, most human rights adjudicative mechanisms issue much more rulings against States than they do against petitioners and they may still be legitimate for both parties–" but necessarily requires equality in arms, due process and consistent application and interpretation of the law.

As to the interpretation and application of the rules consistent with its scope and purpose, this is probably among the key elements to making FB Oversight Board legitimate. Community standards have a scope and a purpose and they daily dialogue with other norms, including national, regional and international human rights norms. The decisions and rationale behind individual case solutions (if FB so chooses) should follow the same logic and dialogue between community standards and human rights norms. Additionally, that logic should be transparent and published. The independent and objective reasoning behind each decision is what an experts' board adds to a closed system like the existing legal team at FB.

Finally, transparency in this context may be defined as a quality: an office or a body, whether judicial or otherwise, is transparent "when it creates the conditions that allow society to fully and clearly understand how they

² Richard H. Fallon, *Legitimacy and the Constitution*, 118 HARV. L. REV. 1787 (2004–2005)

³ Daniel Bodansky, *The Legitimacy of International Governance: A Coming Challenge for International Environmental Law?*, 93 AM. J. INT'L L. 596, 601 (1999).

⁴ N. Grossman, *Legitimacy and International Adjudicative Bodies*, 41 Geo. Wash. Int'l L. Rev. 107 (2009–2010)

act, the reasons behind their acts, as well as the costs and resources associated with those actions.”⁵ This factor affects the previous two as well in that without some degree of transparency (manifested through some or all of the following: publishing decisions or having them publicly available, having reasoned decisions, identifying the decision makers, their dissents and concurrences, etc.) there is no way of evaluating whether a body is biased or not, or whether its decisions and interpretations of the norms are within the reasonable scope and purpose of the law.

While the legitimacy of international adjudicating bodies traditionally derives from the state’s consent to its jurisdiction, the legitimacy of FB Board would derive from: i) building a solid foundation on fair decisions; ii) a consistent and persuasive interpretation --persuasion is one of the legitimacy function-- of the community’s rules in dialogue with international human rights standards; iii) and holistically transparent mechanisms.

Representation vs diversity

One of the main goals of FB Board is bringing greater legitimacy to its content moderation system. Diversity is often mentioned as one dimension that adds towards the legitimacy of a body, particularly when the constituency that the body governs is diverse. Diversity has many meanings though and should not be confused with representation.

Obviously representation could enhance the legitimacy of the Board’s decisions in the eyes of the broader community, particularly those that make it to being represented within the Board. However, a 40-member Board like the one that is being proposed could hardly represent a 2.7 billion member community. It is practically impossible to provide representation for all. Not direct, proportional, or even asymmetrical representation could be accomplished in such an uneven ratio and because of the scale that FB has, there probably couldn’t be a viable ratio to work with. This issue should be acknowledged and incorporated into the design of the Oversight Board so as to create realistic expectations. Ignoring it would probably mislead users, creators and bystanders and will severely undermine the ultimate goal.

Lacking representation, which could provide for diversity of voices, technical expertise could be the next best thing. Not a representative Board but a technical body capable of overseeing FB’s implementation of their own policy, in dialogue with human rights norms. Such technical membership should consider cultural, geographic and gender diversity towards its conformation to allow for an actual dialogue among different experts and avoid a single voice/view approach. This diversity is key to guaranteeing closeness and cultural understanding across borders; deal with novel interjurisdictional issues, and serving the vast community that FB serves. There is no unique model to fit every need. Deciding what FB Board will be necessarily implies defining and accepting what it will not be.

⁵ M. Pulido Jiménez, M. González Armijo, M. Sánchez de Tagle, S. Ruiz Cervantes, J. Sáenz Andujo, “Hacia un modelo de transparencia y acceso a la información en el Sistema Interamericano de Derechos Humanos”, en *Desafíos del Sistema Interamericano de Derechos Humanos. Nuevos tiempos, viejos retos*, capítulo 3. Colección DeJusticia, 2015, p. 112. Disponible en: <https://bit.ly/2hoG7jt>.

What is expected of the Board broadly?

The draft charter suggests that “The board will be a body of independent experts who **will review Facebook’s most challenging content decisions** - focusing on important and disputed cases. Among the first questions that may be raised is what for? What is the ultimate goal in having the Board review these decisions? This question impacts directly on the type of structure that FB is trying to create. Is this a Supreme Court, a Court of appeals, a peoples’ court? Or is it a panel of peers’ approach? Can it be an advisory body or are we set on a reviewing nature? How will this Board interact with other existing structures within Facebook (security and safety, policy and outreach, to name a few?)

If it is in fact a “Supreme Court-like approach”, which seems to be the rationale for it, Yale Professors Klonick and Kadri argue that “what this really means for free speech and fair process on the internet will depend on the answer to one key question: How much will the “Supreme Court of Facebook” be like the Supreme Court of the United States?”⁶ As they argue in their *New York Times* piece, a key element to the US Supreme Court (like every other supreme court) is that it is bound by a set of rules that remain unchanged through time: the Constitution. But, as described in brief background section to this paper, FB policies change every week and so far there has not been a formal adoption of any specific standard or rule to enlighten that process. What should the equivalent be for FB Board?

UN Special Rapporteur David Kaye argues that international human rights norms should be the ultimate rules to govern online content moderation.⁷ The adoption of such standards provides a universal basis and a somewhat common understanding on what free speech means and what guarantees should be considered when limiting it. It would also provide common language to define and understand some restrictions, avoiding contradictory and ever-expanding terms (like hate speech). Ultimately the adoption of universal rules on human rights, particularly free speech, would also guarantee some certainty against discrimination and abuse, whether these arise from governments, users or advocacy groups, and would contribute to ensuring that no one is “a priori” excluded from public debate, which is the standard set by the Inter-American Court of Human Rights in Advisory Opinion 5 of 1985, one of the most progressive and protective international interpretations of freedom of expression and access to information.

Having international human rights norms be the “Constitution equivalent” has many advantages. However, it also has certain limits that cannot be ignored. Key among those is the freedom of the platform to develop and protect its business and to tailor it to different audiences. Example: Adult entertainment is not illegal and is protected under international human rights standards. Following a direct application of international human rights law, no company could prohibit the distribution or uploading of pornography or violent content to its platforms. However, every company must respect and abide by the UN Principle on Business and Human Rights, which creates concrete expectations vis a vis company engagement with human rights, duty to mitigate and prevent violations, duty to provide redress, etc.⁸

⁶ “How to Make Facebook’s ‘Supreme Court’ Work?”, en *The New York Times*, 17/11/18, disponible en: <https://www.nytimes.com/2018/11/17/opinion/facebook-supreme-court-speech.html>.

⁷ <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx>.

⁸ UN Guiding Principles on Business and Human Rights, https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

The dialogue between international free speech standards and Facebook's content rules should be promoted, developed and expanded. Should/could the Board be the body to do that? The answer to this question will probably contribute to defining the goals for this Board.

Should it be an adjudicatory body, what are the models out there and what can be imported to this new structure: arbitral systems; judicial tribunals; international tribunals; media councils; and more.

One of the main goals of FB Oversight Board is to bring greater legitimacy to the content moderation system; but creating a body legitimate enough to do so is among the main challenges.

Although this private board to oversee and unify content moderation decisions is a first of its kind, there are already numerous different models that can and should inform the process that FB is undergoing. First, the company should take advantage of the best practices and lessons learnt from over 100 years of international adjudicatory mechanisms of different sorts if the idea is in fact to create an adjudicative body; second, there is a legitimacy to these bodies that justify their bearing in deciding how to move forward with this particular initiative; third, as opposed to national decision-making bodies, international adjudication was specially designed to deal with cultural and national differences, having gained an expertise on the matter that should be acknowledged and learnt from.

International adjudication mechanisms vary from area to area and from region to region. There are different models that could be looked at, including the International Court of Justice; the different arbitration mechanisms that were created to deal with bilateral investment treaties (i.e. ICSID); the Human Rights Committee created under the International Covenant on Civil and Political Rights; the different regional human rights adjudication mechanisms that exist (i.e. European Court of Human Rights, Inter-American Court of Human Rights, African Commission on Human and Peoples' Rights). These are but a few examples of international adjudicatory bodies that already exist and that have faced already some of the most challenging issues that FB Board will most probably face -i.e. cultural differences, language, nationality, global norms and standards; international sources of law.-

The design chosen for the Board will of course determine a number of other things, including the nature and standing of its members, the dynamics that the Board is expected to have, the way and tools it will have to deal with diversity in all its shapes and sizes (criteria, language, legal culture, etc.). Arbitration panels for example are case specific, diverse, made up of individuals selected from an existing list of accredited arbitrators. They rely upon a strong secretariat to maintain the process on track and guarantee minimum procedural cohesiveness and some institutional memory that is helpful in processes that are often times confidential. The ability to choose arbitrators helps the parties build trust around the process, providing for an opportunity for each party to choose an arbitrator and having the third usually being appointed by the Secretariat. This also contributes to guaranteeing some familiarity between the decision-makers and the issues, context, language and culture that the case has arisen in. The diversity of the pool of arbitrators is particularly relevant to these structures. Still, the decentralized nature of the body itself can attempt against cohesiveness of the decisions arising from the body. The fact that the panel is case specific can be problematic when trying to define through interpretation the meaning and implications of an otherwise general or broad international standard.

International tribunals like the Inter-American Court of Human Rights or the European Court of Human Rights are radically different from arbitration panels. They have a consistent membership through the years, which guarantees some consistency and legal certainty as to the decision-making criteria. It also guarantees a greater level of equality for their users. Most of the times there is cohesiveness to their decisions and their permanent nature provides more transparency and accountability of the panel itself. The downside of these structures is that guaranteeing diversity and representation of every potential party is not possible and they are probably more constrained in the number of cases that they can review in a given year. The Inter-American Court for example always hears cases in full. The European Court has set up a system whereby the members of the Court sit in different chambers, each having 3 judges appointed to them. Some decisions are brought to the Grand Chamber and therefore they guarantee the cohesiveness of their decision-making more broadly.

As set out in the brief examples, there are different potential models that FB could adopt in designing their Board. Still, regardless of the one they choose, comparative and previous experiences should be consulted and taken into account when finalizing this project.

In taking these tribunals and its practices under consideration, particular attention should be paid to the lessons learned and the worst and best practices arising from them. One of the many problems that these tribunals face is the backlog that they generated. Cases take a long time to be litigated before these bodies, whether they require a lawyer or not, and access to international tribunals is not easy. Key among the questions that FB should ask itself is how is this “private quasi tribunal” going to work in scale? Will they be adding to a systematic crisis worldwide *vis a vis* access to justice, or will they be contributing to at least partly solving that issue. The goals and expectations for the Board are not a minor thing to consider here: if this is an appeals Court, a users resort of some sort, how will this body deal with 2.7 billion users and the amount of content they generate?

FB already faces criticism for their internal immediate processes for reviewing company decisions on content moderation. These include not only removal decisions, but also decisions to downgrade or foster the circulation of certain contents versus others. These issues will not be solved with an Oversight Board and in fact, the creation of an Oversight Board should not draw the attention away from them -regular appeals mechanisms within the company-, as these are the basis for any potential redress for wrongful or unfair content moderation practices.

A Supreme Court (or Constitutional or international Court) style approach could make a much more substantive contribution and fit more smoothly into the scaling issue. However, there should be a lot more clarity as to expectations, criteria, process, case selection, standing (for NGOs, users, consumer organizations?), etc. for this board to receive and select cases, deal with them and make them public. Decisions arising from the Board should also impact the resolution of similar cases within FB regular content moderation operations and appeals processes, thereby turning the Board into an internal reference and authoritative body for the company to interpret ToS in dialogue with international human rights law. Otherwise, and because of issues of access, scale and relevance, the exercise will soon derive moot.

Who will be part of this Board? Requirements to be members of the Board?

In thinking about adjudicative bodies, who decides is as important as how they decide. Still, the question that we propose is not literally who will be on the Board but rather what will the criteria be for selecting those mem-

bers. The question is very much linked to the expectations that one may have for this board. Who or what is it overseeing and what for?

Selection criteria is key to guaranteeing legitimacy as explained above. Comparative and existing bodies should also be brought to bare in defining the eligibility criteria for Board members. Knowledge and expertise in international human rights and particularly free speech should be among the qualities of any candidate. This recommendation should not be taken lightly. FB already has security and safety councils where third party experts participate and actively engage in designing the company's policy and terms of service. Reviewing content moderation and curation decisions necessarily implies balancing freedom of expression against other rights.

Most international human rights adjudicative bodies require that its members meet the requirements in their own countries to be judges and that they have demonstrated expertise in human rights. While not every judge will have the same education (some will be from common law countries, others from civil law traditions; some will be from the global north, others from the global south), they will all have some sort of legal education. Members of the Inter-American Commission or the UN Committee need not be lawyers but need a number of years of experience, sound knowledge of international human rights law and high moral and ethical standards. In order to be in an arbitration panel, there are certain requisites including being versed in law that arbitrators need to meet.

As opposed to other structures common to a number of different companies -like the security and trust committee, or the children's safety group- the Oversight Board that is being proposed is intended to deal with limits and restrictions over speech. While these other bodies are mostly made up of experts on children's issues, vulnerable populations, risk management, violence and abuse, there probably are not many free speech experts within those groups. Different areas require different skills and a key question that FB should ask in defining criteria for the Oversight Board is what their role will be and what their skills should be.

Conclusions:

First and foremost, in designing FB Oversight Board, FB and its team should evaluate the impact of such a body on the human rights of its users, particularly freedom of expression, due process, access to justice, equality and non-discrimination. It should also evaluate how this Board, whatever the structure it ends up having, will contribute towards implementing the UN Business and Human Rights Principles.

Legitimacy is a key element that both FB and its users crave in its content moderation. Still, legitimacy may be defined in different and varied ways and legitimacy among decision-making bodies require certain key characteristics that should not be ignored if the goal is set for an adjudicative body. Scale and diversity in this particular case -with a 2.7 billion user constituency across 180 countries- pose additional and complex challenges to accomplishing legitimacy in traditional representational models. These challenges should be acknowledged and addressed systemically. FB should not target that which it cannot provide.

There are still more questions than answers surrounding the creation of FB Oversight Board. In defining and answering those questions, particular attention should be paid to the objectives and expectations for the Oversight Board. After many consultations, it is clear that different organizations and different people will have different expectations for this body. Clarifying what FB is thinking on these points is key and it will be important to provide

new spaces for dialogue and comments after the company lands a concrete proposal for its Board and before it actually implements it.

Should FB decide to create a private adjudication model, it should bare in mind best practices and lessons learned from the many and diverse structures that adjudication mechanisms have adopted over the years, with varying results. International adjudication mechanisms are particularly relevant to look at since they are cross-borders, serve an inter-jurisdictional multicultural constituency, are courts and mechanisms of last resort, and usually resolve complex and intertwined legal issues.

Following FB human rights obligations as well as the particular framework that speech restrictions enjoy internationally, particular attention should be paid to the technical expertise required for members of the Board to join. Developing concrete, specific and clear criteria for the selection of the Board is key to its founding, regardless of who chooses the first set of members. The selection criteria will depend on the objectives and the concrete expectations for this Board.

About CELE

The Center for Studies on Freedom of Expression and Access to Information is a research Center housed at Universidad de Palermo, in Buenos Aires, Argentina. The Center provides legal technical research to promote the understanding and development of freedom of expression and access to information, particularly in Latin America. Since 2012 we have an Initiative for Freedom of Expression online (iLEI- Spanish) under which we have studied and produced research pertaining to online free speech, access to information and privacy particularly under the framework of the Inter-American human rights system and standards. Our strategies to affect change include research; capacity building; and promoting spaces for high level reflection and debate. Please visit us at www.palermo.edu/cele and at www.observatoriolegislativocele.com.