Shedding Light on the Darkness of Content Moderation

The First Decisions of the Facebook Oversight Board

Oreste Pollicino & Giovanni De Gregorio

05.02.2021 doi: 10.17176/20210205-235959-0

About two years ago, Mark Zuckerberg described his vision on how online content should be governed on Facebook. "Facebook should not make so many important decisions about free expression and safety on our own". He imagined a new system, in which users could rely on an independent body overseeing decisions on content moderation. The recent establishment the Facebook Oversight Board (FOB) has manifested this idea, and its first decisions promise to provide users with a new potential remedy against discretionary content removal, while making Facebook's decision-making on content moderation more accountable and transparent.

With the FOB, we face a new age of private adjudication of online content, which promises an alternative system to enforce human rights on a global scale, while marginalising and hybridising constitutional values and democratic safeguards. Digital constitutionalism offers a framework to look at this new form of private adjudication of online content and its challenges. This is no abstract deliberation: the genocide in Myanmar, the recent events in the United States and the pandemic have highlighted how private ordering by social media has an impact on individual rights and democracy.

In other words, the FOB's first cases are not only an opportunity peek behind the scenes of content moderation choices, but also a laboratory to study the transnational challenges which the information society has raised to global (digital) constitutionalism.

The Cases in Short

On 28 January, the FOB published its first set of decisions covering four continents and issuing nine policy recommendations. The five available decisions review Facebooks' determinations via a mix of community guidelines and international human rights law. Most of them were unanimous. Just one of them upheld Facebook's original decision, while the others overturned removals.

The FOB overturned Facebook's decision to remove content shared by a user in Myanmar disseminating expressions of hate against Muslims. This decision was based on the translation and recontextualization of the content published by the user, which meant to compare extremist religious responses in different countries. Likewise, contextualization was also relevant in another decision on hate speech. In this case, the FOB overturned a removal of content which quoted Joseph Goebbels, to underline that arguments appealing to emotions and instincts should prevail over truth. The FOB gave precedence to the protection of freedom of expression, since the user was not supporting Nazi propaganda but commenting on Donald Trump's actions.

In a decision on health disinformation, freedom of expression also prevailed. A user argued in favour of Hydroxychloroquine and Azithromycin for the treatment of COVID-19 and complained about the lack of authorization by the French agency responsible for regulating health products. In this case, by adopting a proportionality-driven approach, the FOB concluded that this statement could only cause minimal harm, *inter alia,* because these drugs require a prescription in France.

The FOB also addressed a Brazilian case where content was removed without human oversight. Photographs of breast cancer symptoms, to raise awareness of the disease, were removed automatically from Instagram for violating community standards on adult nudity and sexual activity. Here, the failure of artificial intelligence systems to recognize Portuguese expres-

sion led to a mistake which, according to the FOB, requires that Facebook specifies that such content can be published for awareness-raising purposes.

The only case in which the FOB upheld Facebook's decision concerned hate speech against Azerbaijanis, who were called *taziks*, which is considered "a dehumanizing slur attacking national origin". In this case, taking into consideration the violence between the two countries, the FOB gave prevalence to values such as dignity and safety according to international human rights standards, over the protection of speech.

Enforcing Human Rights in the Darkness of Private Standards

The challenges and opportunities of this new system of private adjudication are well–discussed: The consensus is that the relevance of the FOB's decisions should neither be underestimated, nor overestimated. These cases do not only promise more responsibility and accountability of Facebook's governance of online speech, but also showcase private actors' power to autonomously enforce community guidelines, while seeking legitimacy from an (independent) Oversight Board.

While the FOB's reasoning is based on international human rights law, it is inevitably also influenced by private standards defined in Facebook's community guidelines. Facebook and other social media usually rely on terms of service and community guidelines in which they set standards of acceptable conduct and content. In other words, these private rulebooks form a quasi-legal basis according to which platforms exercise their powers, sans reliance on public authorities. Via these quasi-normative and executive powers, online platforms perform functions which are similar to that of the judiciary, and especially of constitutional courts: the balancing of fundamental rights. In other words, the FOB is another step towards private enforcement, and the institutionalisation of private adjudication.

From a constitutional perspective, Facebook's choices and actions represent a form of self-regulation and marginalisation of the role of public actors in ensuring the enforcement of fundamental rights online. Social media as private actors are not bound by human/fundamental rights, absent regulation requiring them to translate constitutional values into their business activities. Even if the FOB were to improve transparency and accountability in Facebook's content moderation, their handful of cases are a small drop in the ocean of content which is moderated against opaque self-set standards every day.

The Decline of Juristocracy?

Enthusiasm and scepticism characterize the debate around the FOB, which has been called the "Facebook Supreme Court". However, its decisions are not only taken outside democratic oversight, but do also not reflect the traditional safeguards guaranteed by constitutional principles like the rule of law, due process and equality of arms. The FOB is not a real court. For this reason, these first decisions cannot be compared with either *Marbury v. Madison*or *West v. Barnes*in the US, or with*Lawless v. Ireland* in Europe.

Rather, the FOB is more a quasi-court, acting as a check on a private platform defining standards of speech protection on a global scale, while playing an advisory role like the European Court of Human Rights after the French ratification of Protocol 16 in 2018. The FOB is itself a private institution, whose legitimation and resources depend on its private founder. Even though Facebook has invested $130 million into a trust to support the Board's activities, Facebook still plays a critical role in influencing its oversight. For instance, Facebook participates in the selection of cases to refer to the Board. Further, the FOB's "jurisdiction" derives from what Facebook decides is in scope. It's limited to cases concerning content removal or suspension, without the possibility to assess Facebook's decisions to keep content online. Reviews only concern the phase of 'hard' moderation (e.g. content removal), not 'soft' moderation (e.g. recommending). Facebook has an interest in shielding 'soft' moderation from FOB scrutiny, since the delicate cases involved (e.g. political advertising) are integral to the platform's business model.

The challenges of private adjudication fall within the transnational trend in which states struggle to maintain their role, while ceding powers to fourth (e.g.

national and supranational agencies) and even fifth (private) branches. Post-modern constitutionalism is characterised by a process of fragmentation in terms of sovereignty, followed by a parallel process of reconfiguration within a multilevel and polycentric system. Therefore, it is essential to establish the correct and fastest interconnections between the different constitutional centres that frame the new polycentric global order at national, supranational and international level, including private ordering. In the current era of legal and economic globalisation, it is worth observing that traditional constitutional governance is altering the characteristics that marked the process of its development in previous centuries. In particular, it is possible to observe a definitive decline in the historical constellation that featured the simultaneous presence, within the same national borders, of state, sovereignty and economy triangle.

Within this framework, it is worth considering the future of juristocracy – one of the primary trends within post-modern constitutionalism in the era of judicial globalisation. Global governance seems to prefer the language of the 'law in action' rather than the ink of the 'law in the books'. The FOB perfectly fits within this framework. Still the primary question is whether this governance will come from courts or private adjudication bodies. It is likely that courts will by no means lose the predominant role over political power acquired in recent years. But how they will deal with alternative models of adjudication is still to say. The answer to this question is far beyond the analysis of the FOB decisions but the consolidation of these areas of private *ius dicere* constitutes a transformative challenge for the role of courts in the information society.

Welcome to the New Age of Private Adjudication

We cannot predict the impact of the FOB's decisions in the long run. In the interim, as the first decisions show, the FOB may increase transparency in content moderation, while making Facebook more accountable for its choices. Nonetheless, we should not underestimate the challenges of omissively delegating private actors quasi-regulatory powers over users' rights and freedoms. This is a primary challenge for digital constitutionalism. The original mission of constitutionalism is to set some mechanisms to restrict government power through self-binding principles, including different forms of separation of powers and constitutional review. The FOB has rung in a new phase of private adjudication, raising questions about the role of constitutional law in limiting private ordering which impacts fundamental rights' and democratic values.

The FOB will not solve the primary challenges of content moderation like advertising and data profiling, outsourcing of human moderator, artificial intelligence biases, or generally, the global impact of online speech. Instead, this new model of private adjudication could charm and reassure public actors about the virtues of self-regulation, while consolidating areas of powers escaping democratic oversight. The consequence of this process would affect juristocracy. Not only could the FOB lead to a marginalisation of judicial power but also contaminate judicial values and safeguards which would be driven by private ordering on a global scale.

Against this trend, in Europe, we are experiencing a new path toward mitigating social media powers by proceduralising content moderation. The Digital Services Act (DSA) proposal is just an example of the path of European digital constitutionalism with the goal of injecting due process safeguards in the moderation activities of online platforms. Nevertheless, before the DSA enters into force, the primary challenge is to understand the role of constitutional values in avoiding FOB-like laboratories of private adjudication.

ABOUT THE AUTHORS

Oreste Pollicino
Oreste Pollicino is Professor of Constitutional Law at Bocconi University in Milan, where he also teaches Information and Internet Law, Public Law and Transnational Constitutional Law. He is a member of the Managing Board of the European Agency for Fundamental Rights.
Giovanni De Gregorio
Giovanni De Gregorio holds the PLMJ Chair in Law and Technology at the Católica Global School of Law, Lisbon.Giovanni De Gregorio holds the PLMJ Chair in Law and Technology at the Católica Global School of Law, Lisbon.

SUGGESTED CITATION

Pollicino, Oreste; De Gregorio, Giovanni: Shedding Light on the Darkness of Content Moderation: The First Decisions of the Facebook Oversight Board, Verfassungsblog, 05.02.2021, doi: 10.17176/20210205-235959-0.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20210205-235959-0.