Keele Law Review, Volume 3 (2021), 90-99
ISSN 2732-5679

Moderating Misinformation During the COVID-19 Pandemic: Why Social Media Platforms Need an Emergency Constitution

João Victor Archegas*

I. Introduction

How should social media platforms regulate speech during a public health emergency? During the COVID-19 pandemic, it became common sense to argue that Facebook and Twitter need to work in harmony with public health officials to curb the spread of harmful posts, especially user-generated content that may undermine the implementation of safety protocols (e.g., social distancing and face masks). But to do that in a reasonable, transparent, and predictable way, these platforms should incorporate an "emergency constitution" to their community standards. In other words, Facebook and Twitter, just like different national governments around the world, should have the option of exercising emergency powers, which would allow for a more efficient and accountable content moderation when circumstances call for quick responses from our tech overlords.

In the US, platforms enjoy a broad protection from liability for content posted by their users. Section 230 of the Communications Decency Act of 1996 provides that "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider". The Section 230 immunity was created by Congress in the 90s as a response to two relevant court cases that were decided in 1991 and 1995.

The first case is Cubby. v. CompuServe (1991), in which the United States District Court for the Southern District of New York held that CompuServe was a distributor and not a publisher of content and, therefore, could only be held liable if it knew or should have known of any illegality that was taking place on its platform.[1] The second case is Stratton Oakmont v Prodigy Services (1995), in which the New York Supreme Court held that Prodigy Services could be held liable for the speech of its users because it was actively moderating content that was uploaded to the platform.[2]

Section 230 solves the tension created by these two cases by stating that online service providers like CompuServe and Prodigy Services cannot be treated as publishers or speakers of information hosted on their platforms, even if they actively moderate user-generated content. Under Section 230 of the CDA, tech companies are not obliged to screen their platforms for offensive or illegal content. Nevertheless, if they opt to do it (like Prodigy Services did), they will not be held liable in case they fail to remove illegal

* João Victor Archegas holds an LLM from Harvard Law School and is currently a researcher and professor at the Institute for Technology and Society of Rio de Janeiro, Brazil. The author would like to thank the participants of a graduate workshop at the Federal University of Paraná, Brazil, for all their thoughtful comments and insights. João can be reached at j.archegas@itsrio.org.

[1] Cubby, Inc. v. CompuServe, Inc.,(1991) 776 F.Supp. 135.

[2] Stratton Oakmont, Inc. v Prodigy Services Co.,(1995) 23 Media L Rep 1794.

content. Similar provisions exist in other relevant jurisdictions, including the European Union.[3]

But if Facebook and Twitter do not have to fear liability for third-party content, why bother moderating content at all? According to Kate Klonick, there are three main factors behind the development of moderation systems by these platforms.[4] The first one has to do with "free speech norms". Their standards were originally developed by American lawyers who take the First Amendment to the US Constitution as the baseline for speech moderation. Second, platforms have an incentive to protect their users' speech from undue government intrusion. Public officials are always pushing for more content moderation, which can lead to growing censorship creep on the Internet.[5] Finally, economic incentives also play a determinant role. Digital platforms wish to create a friendly environment to attract more users and promote engagement, which will often entail the removal of offensive content.

Although self-regulation was the norm for many years within the industry, recently the number of government regulations began to rise. The trend was prompted by a wave of terrorist attacks in Europe and other parts of the world. After the 2019 terrorist attack in Christchurch was broadcasted live on Facebook, New Zealand and France unveiled the Christchurch Call to Action, a set of "collective, voluntary commitments from Governments and online service providers intended to address the issue of terrorist and violent extremist content online and to prevent the abuse of the internet as occurred in and after the Christchurch attacks."[6]

Other governments have been more aggressive, passing new laws or unveiling new regulatory schemes that can be implemented in the future. Germany, for example, enacted the *Netzwerkdurchsetzungsgesetz* (NetzDG). Under the German law, social media companies are subject to fines of up to 50 million euros if they fail to remove "obviously illegal" content hosted on their platforms within 24 hours of it being reported.[7] On the other hand, the UK published the Online Harms White Paper in April of 2019 (now known as the Online Safety Bill). The document outlines a "new system of accountability and oversight for tech companies" and is designed to move beyond self-regulation by creating an independent regulator that will be responsible for enforcing a new statutory duty of care.[8]

Despite the recent rise in government regulation in the last few years, self-regulation can find a lifeline in the model advanced by the Facebook Oversight Board. Initially named the Facebook "Supreme Court" by CEO Mark Zuckerberg, the Oversight Board "will be in

---

[3] Giancarlo Frosio (ed.), *The Oxford Handbook of Online Intermediary Liability* (OUP 2020).
[4] Kate Klonick, 'The New Governors: The people, rules, and processes governing online speech'(131 HarvLRev 2018), pp. 1616-30.
[5] Danielle Citron, 'What to do about the emerging threat of censorship creep on the Internet' (2017), Cato Institute Policy Analysis No. 828.
[6] 'Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online' (15 May 2019). <https://www.christchurchcall.com/christchurch-call.pdf> accessed 28/02/2021.
[7] Center for Democracy & Technology, 'Overview of the NetzDG Network Enforcement Law' (17 July 2017). <https://cdt.org/insights/overview-of-the-netzdg-network-enforcement-law/> accessed 28/02/2021.
[8] UK Secretary of State for Digital, Culture, Media & Sport and UK Secretary of State for the Home Department, 'Online Harms White Paper' (December 2020). <https://www.gov.uk/government/consultations/online-harms-white-paper> accessed 28/02/2021.

charge of adjudicating appeals from users whose content has been removed from Facebook's platforms. It will also make judgements on cases referred to it by the company itself".[9] With that move, the company wants to outsource part of the responsibility that comes with self-regulation to an independent board that will issue final and binding decisions on online speech. The Board delivered its first decisions in the beginning of 2021 and, just a few months later, decided that Donald J. Trump was correctly deplatformed by Facebook following the invasion of the Capitol on January 6.

According to Evelyn Douek, there are at least four reasons for having an independent Oversight Board with the power of issuing binding decisions on content moderation.[10] First, it can bestow legitimacy upon the platform and reassure users. Second, by delegating decision-making powers to an independent body, Facebook can dodge external regulation and have a say in the regulatory framework that it will need to comply with. After all, the Board has the power to not only decide on the merits of a given content moderation case (i.e., whether a post was rightly removed for violating Facebook's hate speech policy), but it can also issue policy recommendations. Third, Facebook will be able to outsource controversial decisions. This has long been Zuckerberg's ultimate wish, who is uncomfortable with the idea of serving as an "arbiter of truth". Finally, the board can also enforce Facebook's existing policies in a more rational and efficient fashion.

In this essay, my objective is to show how Facebook and Twitter (with a focus on the former) are dealing with content moderation during the COVID-19 pandemic and what can be done to make this process more reasonable, predictable, and transparent. My proposal is that digital platforms should take a cue or two from the "emergency constitution" and rewrite their community standards to authorize the use of emergency powers during public emergencies. In the next section, I explore some of the steps that these companies have taken to moderate harmful COVID-19 misinformation. I then assess changes in their policies in light of digital constitutionalism. In the third section, I explore the literature on comparative emergency powers and argue that digital platforms are operating under a *de facto* state of emergency. In the fourth section, I make some recommendations that should be implemented by social media platforms. Finally, in the fifth section, I offer a brief analysis of how the Oversight Board is already grappling with some of the questions raised by this essay.

II. Speech Regulation and Digital Constitutionalism

As argued above, the regulation of online speech is done mostly by online service providers like Facebook and Twitter through their community standards, a set of written rules that determine the boundaries of free speech online. Although some countries passed new statutes to regulate online speech in the last few years, intra-industry initiatives like Facebook's Oversight Board show that tech companies are not going to give up on their regulatory autonomy so easily. For better or worse, tech behemoths will have a say on speech regulation for years to come.

---

[9] Makena Kelly, 'Facebook's 'Supreme Court' can overrule Zuckerberg, per new charter' (The Verge, 17 September 2019). <https://www.theverge.com/2019/9/17/20870827/facebook-supreme-court-mark-zuckerberg-content-moderation-charter> accessed 28/02/2021.

[10] Evelyn Douek, 'Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility' (2019) 21(1) North Carolina Journal of Law & Technology, pp. 15-28.

Thus, the question is not whether we should allow Facebook to self-regulate. Instead, the question should be whether Facebook is doing everything it can to curb online misinformation, especially "fake news" related to the novel coronavirus pandemic. On April 16, 2020, the company announced that it would redirect users that have engaged with "harmful coronavirus misinformation" on its platform to the "myth busters" page curated by the World Health Organization (WHO).[11] Furthermore, Facebook "started removing claims that physical distancing doesn't help prevent the spread of the coronavirus" and "banned ads and commerce listings that imply a product or guarantees a cure or prevents people from contracting COVID-19."[12]

Before announcing this new framework for dealing with coronavirus disinformation, CEO Mark Zuckerberg said that "even in the most free expression-friendly jurisdictions like the US, you've long had the precedent that you don't let people yell 'fire' in a crowded room, and I think that's similar to people spreading dangerous misinformation in a situation like this".[13] Even Twitter, which has historically resisted calls to moderate speech, announced that it would revise its internal rules, broaden its definition of harm "to address content that goes directly against guidance from authoritative sources" and remove tweets that include "denial of [...] health authority recommendations" and "description of harmful treatments or protection measures which are known to be ineffective".[14]

Acting in accordance with these new community standards, both companies removed world leader's posts from their platforms.[15] This is a rare decision coming from businesses that have resisted calls to censor world leaders. The example coming from Brazil is notable. President Jair Bolsonaro is a far-right politician who has been downplaying the crisis ever since the first case of COVID-19 was confirmed in the country. According to the Washington Post, that makes him one out of four heads of state who are downplaying COVID-19 and arguably the worst among them.[16] Bolsonaro infamously called the novel

---

[11] Guy Rosen, 'An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19' (Facebook Newsroom, 16 April 2020). <https://about.fb.com/news/2020/04/covid-19-misinfo-update/> accessed 28 February 2021.

[12] Nick Clegg, 'Combating COVID-19 Misinformation Across Our Apps' (Facebook Newsroom, 25 March 2020). <https://about.fb.com/news/2020/03/combating-covid-19-misinformation/> accessed 28 February 2021.

[13] Alex Hern, 'Covid-19 outbreak like a nuclear explosion, says archbishop of Canterbury – as it happened', (The Guardian, 18 March 2020). <https://www.theguardian.com/world/live/2020/mar/18/coronavirus-live-news-updates-outbreak-us-states-uk-australia-europe-eu-self-isolation-lockdown-latest-update?page=with:block-5e727a0a8f088d7575595fd9#block-5e727a0a8f088d7575595fd9> accessed 28 February 2021.

[14] Matt Darella, 'An update on our continuity strategy during COVID-19' (Twitter Blog, 16 March 2020). <https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html> accessed 28 February 2021.

[15] BBC News, 'Coronavirus: World leaders' posts deleted over fake news' (31 March 2020). <https://www.bbc.com/news/technology-52106321> accessed 28 February 2021.

[16] Editorial Board, 'Leaders risk lives by minimizing the coronavirus. Bolsonaro is the worst' (Washington Post, 14 April 2020). <https://www.washingtonpost.com/opinions/global-opinions/jair-bolsonaro-risks-lives-by-minimizing-the-coronavirus-pandemic/2020/04/13/6356a9be-7da6-11ea-9040-68981f488eed_story.html> accessed 28 February 2021.

coronavirus "a little flu" and has urged his supporters to ignore regional and local lockdown orders.[17]

Facebook and Twitter deleted two videos that Bolsonaro uploaded to both platforms using his official, presidential profiles. The first one showed him claiming that hydroxychloroquine is a sound and effective treatment for COVID-19, even though medical evidence is still inconclusive at best. In a second video, Bolsonaro is shown greeting supporters outside the presidential residence and causing agglomerations in Brasília, the country's capital. The videos were flagged for "causing harm" – which is a violation of the community guidelines on both sites – and were duly deleted from the platforms.

More recently, on February 8, 2021, Facebook announced that the company would outright ban false or misleading statements about vaccines from the platform, including the age-old anti-vax argument that vaccines cause autism.[18] Although the new policy was prompted by concerns over COVID-19 vaccination across the globe, the new prohibition extends to all vaccines and covers the anti-vax movement at large. Two days after announcing the change, Facebook banned Robert F. Kennedy, a well-known anti-vaxxer, for "repeatedly sharing debunked claims about the coronavirus or vaccines".[19] Similarly, on a blog post from December of 2020, Twitter announced that it would expand its policy to "require people to remove Tweets which advance harmful false or misleading narratives about COVID-19 vaccination".[20]

The problem is that the new standards for regulating speech during the pandemic seem to come out of thin air. Social media platforms have long been accused of censoring speech without clearly exposing and articulating the reasons behind their controversial decisions. After the Guardian leaked Facebook's policies for content moderation in 2017,[21] the company responded by making its community standards public for the first time and introducing an appeals process.[22] This was definitely a step in the right direction, but users remain suspicious about the reasons driving the company's decision-making process. This suspicion only grows when tech companies are constantly changing their policies around COVID-19 and vaccine disinformation through blog posts and press releases that fail to shed light on their internal procedures. Constant changes coupled with

---

[17] Simone Iglesias *et al*, "Little Flu' Can't Hurt Him: Why Bolsonaro Still Shuns Lockdowns' (Bloomberg, 30 March 2020). <https://www.bloomberg.com/news/articles/2020-03-30/-little-flu-can-t-hurt-him-why-bolsonaro-still-shuns-lockdowns> accessed 28 February 2021.

[18] Mike Isaac, 'Facebook says it plans to remove posts with false vaccine claims' (New York Times, 8 February 2020). <https://www.nytimes.com/2021/02/08/technology/facebook-vaccine-misinformation.html> accessed 28 February 2021.

[19] Jon Porter, 'Instagram bans prominent anti-vaxxer Robert F. Kennedy, but Facebook page remains active' (The Verge, 11 February 2021). <https://www.theverge.com/2021/2/11/22277880/robert-f-kennedy-jr-instagram-banned-covid-19-vaccine-misinformation> accessed 28 February 2021.

[20] Twitter Safety, 'COVID-19: Our approach to misleading vaccine information' (16 December 2020). <https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine.html> accessed 28 February 2021.

[21] Nick Hopkins, 'Revealed: Facebook's internal rulebook on sex, terrorism and violence' (The Guardian, 21 May 2017). <https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence> accessed 28 February 2021.

[22] Casey Newton, 'Facebook makes its community guidelines public and introduces an appeals process' (The Verge, 24 April 2018). <https://www.theverge.com/2018/4/24/17270910/facebook-community-guidelines-appeals-process> accessed 28 February 2021.

lack of transparency are at odds with the values and principles of digital constitutionalism, especially the rule of law.

Digital constitutionalism is "the ideology which aims to establish and to ensure the existence of a normative framework for the protection of fundamental rights and the balancing of powers in the digital environment".[23] In positive terms, it is also about enabling a well-functioning architecture for meaningful communication and connection. Contemporary constitutionalism has a long history, but it basically strives to achieve three objectives. First, it offers a framework for balancing powers within government. Second, it implements an agenda to protect and advance fundamental rights. Third, it enables a functional government. Consequently, digital constitutionalism is an attempt at abstracting the ideals of constitutionalism from the context of the nation-state so they can be reimagined in (and repurposed for) a new setting.[24] To put it differently, it is about translating constitutional principles for the digital realm.

One of the founding principles of constitutionalism is the rule of law. As a matter of principle, any rule-based system must follow a set of basic standards that include, among others, predictability and stability. These are what Lon Fuller calls "principles of legality".[25] To put it simply, because rules are set to constrict and channel specific, predetermined behaviours (i.e., speech), it is of utmost importance that the subject of those rules understand, in advance, what they prescribe. Furthermore, it also matters that rules remain relatively stable over time to foster confidence and trust in the system. Unfortunately, Twitter's and Facebook's community standards run afoul of the rule of law and, consequently, fail to uphold the principles and values of digital constitutionalism.

III. A Digital, *De Facto* State of Emergency

Facebook and Twitter are basically exercising emergency powers to regulate COVID-19 misinformation. Both companies revised their community standards to deal with false stories more rigorously during the public health crisis, removing content that may harm their users. As mentioned above, social media companies frequently announced changes and tweaks to their policies to deal with false claims about COVID-19. But this is alarming for two reasons.

First, there is no clear distinction between the *norm* (the state of affairs prior to the emergency) and the *exception* (the state of emergency itself) on social media platforms. The *norm* is what one may call "regular government". Under a state of emergency, the *norm* is derogated so actions that would otherwise be prohibited or limited can now be performed to cope with the emergency.[26] In other words, social media users are not aware that these platforms are operating under a *de facto* state of emergency because there is no regulatory framework to serve as a baseline.

This is why constitutions around the world provide for the declaration of a state of emergency when some enumerated circumstances or conditions arise, such as a public

---

[23] Edoardo Celeste, 'Digital Constitutionalism: a new systematic theorisation' (2019) 33(1) International Review of Law, Computers & Technology, p. 88.
[24] *Ibidem*, p. 89.
[25] Lon L. Fuller, *The Morality of the Law* (YUP 1969).
[26] Oren Gross, 'Constitutions and Emergency Regimes' in Tom Ginsburg and Rosalind Dixon (coord.), (EEP 2011) *Comparative Constitutional Law*, p. 334.

health emergency or a natural catastrophe.[27] To deal with the novel coronavirus pandemic, a number of countries declared states of emergency, allowing local governments to exercise emergency powers.[28]

Social media platforms are also operating under a state of emergency of sorts, yet the boundaries of their emergency powers are not clearly set. Unlike national constitutions, their community standards do not provide for the derogation of the *norm*. There can be no accountability or predictability when social media platforms can simply declare an emergency by fiat and decide what their emergency powers will look like once the emergency is already underway. Facebook and Twitter need to incorporate a basic set of provisions that define under what circumstances an emergency can be declared and what powers will be available.

Second, emergency constitutions – or the set of emergency provisions enshrined in national constitutions – serve two different functions. First, they allocate to different actors the powers to declare and end an emergency. Usually, the head of state is responsible for declaring an emergency while the legislature has the prerogative of approving or disapproving the declaration. Even if the legislature approves it, the constitution establishes the precise duration of the emergency. Once the declaration expires, the legislature will vote on whether or not it should be renewed for the same period of time. Second, an emergency constitution is designed to prevent the *exception* from entrenching itself as the new *norm*. In other words, emergency powers cannot be used to alter the constitutional framework during an emergency.

According to Ferejohn and Pasquino, "emergency powers in modern constitutions are to be employed to deal with temporary situations and are aimed at restoring the conditions to a state in which the ordinary constitutional system of rights and procedures can resume operation. Typically, the holder of emergency powers is not permitted to make law but is restricted to issuing temporary decrees. And of course, the constitution itself is not to be changed in such periods."[29] To put it differently, emergency powers have two defining features: they are *temporary* – in the sense that they are not part of a regular government and can only be used in exceptional circumstances – and *conservative* – in the sense that they can only be used to restore prior conditions.

Although social media platforms are using exceptional means to moderate speech on their platforms, their community standards offer no answers to some important questions; who inside the company can declare the emergency? Should this decision be subject to review by an independent body? For how long can the emergency endure? Once the original declaration expires, can it be renewed? Who can exercise emergency powers once the declaration is approved? Who gets to monitor the use of emergency powers throughout the duration of the emergency? These, among others, are questions that can only be answered *ex ante* and should be clearly defined in the community standards of

---

[27] Christian Bjørnskov and Stefan Voigt, 'The Architecture of Emergency Constitutions' (2018) 16(1) International Journal of Constitutional Law, p. 101 (according to the authors, "some 90 percent of all constitutions worldwide contain explicit provisions for how to deal with states of emergency").

[28] For an empirical overview of emergency powers during the COVID-19 pandemic, *see* Tom Ginsburg and Mila Versteeg, 'The Bound Executive: Emergency Powers During the Pandemic' (2020) Virginia Public Law and Legal Theory Research Paper No. 2020-52.

[29] John Ferejohn and Pasquale Pasquino, 'The Law of the Exception: A Typology of Emergency Powers' (2004) ICON, p. 212.

social media platforms in order to uphold the principles and values of digital constitutionalism.

IV. Charting a Way Forward

In this section, following the discussion above on the importance of social media platforms adopting an emergency constitution, I offer a few suggestions that can be implemented by Facebook and, with some adaptations, Twitter to deal with emergencies with more transparency and predictability – or in a way that advances rather than violates digital constitutionalism, especially the rule of law. The idea is to foster legitimacy by rationalizing the company's response to the pandemic and empowering users to hold moderators accountable for their decisions during a moment of crisis.

First and foremost, Facebook should consider adopting an emergency constitution as a new and distinct section of the community standards. Facebook's emergency constitution should define *ex ante* what are the circumstances under which the company can declare a state of emergency and, at the same time, enumerate the emergency powers that will be granted to content moderators. It is understandable that the company cannot foresee all the circumstances that may need to be addressed by a declaration of emergency, but it should strive to anticipate the biggest challenges it will face in various crises. For example, during a pandemic or an ordinary public health crisis, the company can declare an emergency to swiftly respond to the spread of content that is antithetical to the recommendations of public health authorities, just like we witnessed in the case of COVID-19 and the WHO.

Second, instead of changing the standards once the emergency is underway, Facebook's emergency constitution would provide for the *temporary* suspension of rights and liberties online. Once the emergency is over, the declaration would be lifted and the community standards restored without permanent modifications. This would prevent the entrenchment of emergency powers and address fears of censorship creep. As I mentioned above, Facebook resisted calls to remove world leaders' posts before, even when it was possible to show that they were completely false. Under a state of emergency, Facebook would be authorized to remove those posts to prevent harm. Once the emergency is over, the company can go back to its deferential stance towards world leaders and resume the protection of free speech on the platform.

Finally, Facebook's emergency constitution would establish procedural constraints on the declaration of a state of emergency. My suggestion is that, once the company declares an emergency, the Oversight Board will be responsible for approving the declaration within 48 hours. If the board approves it by a majority vote, the declaration would be valid for 30 days and could be renewed once for the same period of time. If the declaration is not renewed by the Board after 30 days, the state of emergency will be automatically lifted and the community standards restored. Any attempt at changing the standards during the emergency should be repelled. Furthermore, each step of this process must be made public so users are aware of an eventual declaration of emergency and what that means to their rights and liberties online, especially the right to free speech.

V. The Oversight Board Steps In

In January 2021, the Facebook Oversight Board announced its first five decisions, four of which resulted in the Board overturning Facebook's initial content moderation

judgements. In a sense, this first batch of cases sent a clear message to Facebook; the Board was proclaiming its independence from the company early on. To be sure, the institution was intentionally designed to be as detached from its creator as possible. With that in mind, in this last section of the essay I will evaluate one of these five initial cases to show how the Board is already grappling with some of the questions raised by this essay.

In October 2020, a Facebook user in France uploaded a video to a public group on the platform. The content, which was also accompanied by a text, claimed that the *Agence Nationale de Securité du Médicament* in France should authorize the use of hydroxychloroquine and azithromycin as a treatment for COVID-19. Facebook removed the video asserting a violation of its rule against "misinformation and imminent harm". As the company's reasoning goes, claiming that an untested drug can "cure" the disease may lead people to self-medicate and ignore safety protocols (e.g., social distancing) which, in turn, may cause severe offline harm, especially when the world is facing a deadly pandemic. However, the case was not appealed to the Board by the user. Instead, Facebook independently referred it to the Board "as an example of the challenges of addressing the risk of offline harm that can be caused by misinformation about the COVID-19 pandemic".[30]

Although the Board acknowledged the risk that health misinformation presents, it came to the conclusion that this particular case did not rise to the level of "imminent harm" as required by Facebook's community standards. According to the decision, the company failed to "explain how the post related to imminent harm; it merely asserted imminent harm to justify removal". Furthermore, the Board argued that the user was just advocating for a policy change and not directly encouraging people to self-medicate or disregard safety protocols. Even if the content was interpreted as an encouragement, the Board discovered that hydroxychloroquine and azithromycin are not available in France without a medical prescription. Therefore, if people feel encouraged to use the drugs after watching the video, it is very unlikely that they will have access to the medication. In light of the evidence, the Board decided that Facebook's decision failed the proportionality test.

One of the most interesting things about the decision is that the Board also took issue with Facebook's policy changes during the pandemic, even though that was not necessarily under discussion given the nature of the case. When assessing the legality of the "misinformation and imminent harm" rule, the Board noted that "Facebook has announced multiple COVID-19 policy changes through its Newsroom without reflecting those changes in the current Community Standards. Unfortunately, the Newsroom announcements sometimes appear to contradict the text of the Community Standards". In a strong rebuttal, the Board concluded:

> Given this patchwork of rules and policies that appear on different parts of Facebook's website, the lack of a definition of key terms such as "misinformation", and the differing standards relating to whether the post "could contribute" or actually contributes to imminent harm, it is difficult for users to understand what content is prohibited. The Board finds the rule applied in this case was inappropriately vague. The legality test is therefore not met.

[30] Oversight Board, 'Case Decision 2020-006-FB-FBR' (January 2021). <oversightboard.com/decision/FB-XWJQBU9A> accessed 28 February 2021.

VI. Conclusion

In light of this conclusion and exercising its prerogative of issuing non-binding policy advisory statements, the Board gave two recommendations to the company that are of interest to this study. First, the Board said that "Facebook should clarify its Community Standards with respect to health misinformation". Second, the Board stressed that "Facebook should increase transparency of its content moderations of health misinformation". Although the Oversight Board stopped short of mentioning concepts like "state of emergency" or "emergency powers", it did acknowledge that what Facebook did with its COVID-19 policies is *sui generis* and does not meet the legality test, which is one of the many facets of the rule of law and, likewise, digital constitutionalism. In sum, the Board correctly identified the problem, yet it missed the opportunity to suggest a more specific fix.

As it was argued above, the best way to address problems of legality during a crisis is to adopt an emergency constitution. Declaring an emergency and temporarily suspending the community standards would amount to more reasonable, transparent, and accountable content moderation practices. Besides, it would also reduce the probability of the exception entrenching itself as the new norm online. After the decision, Facebook announced changes to its "misinformation and imminent harm" rule to address the points raised by the Board. According to the New York Times, "the company said the changes were in response to a recent ruling from the Oversight Board" who "said that Facebook needed to create a new standard for health-related misinformation because its current rules were 'inappropriately vague'."[31]

An even more promising step was taken by the Board a few months later, on May 5, 2021, when it upheld Facebook's decision to suspend then-President Donald J. Trump from the platform in the wake of the insurrection that took place in the Capitol on January 6, 2021. By the very end of the decision, in the last paragraph of its policy advisory statement, "the Board urges Facebook to develop and publish a policy that governs its response to crises or novel situations where its regular processes would not prevent or avoid harm."[32] In other words, the Board is (even if only timidly) showing Facebook the importance of adopting an emergency constitution. This is definitely a step in the right direction, but it remains to be seen if Facebook's future interactions with the Board will lead to the adoption of a full-fledged emergency constitution or if the company will keep improvising whenever push comes to shove.

---

[31] Mike Isaac, 'Facebook says it plans to remove posts with false vaccine claims' (New York Times, 8 February 2020). <https://www.nytimes.com/2021/02/08/technology/facebook-vaccine-misinformation.html> accessed 28 February 2021.

[32] Oversight Board, 'Case Decision 2021-001-FB-FBR' (May 2021). <https://oversightboard.com/decision/FB-691QAMHJ/> accessed 24 May 2021.