MEI
Policy Center

HOW FACEBOOK'S OVERSIGHT BOARD CAN DO MORE FOR SYRIA

LEO HOCHBERG

May 2021

Around the world, there is ongoing debate over the extent to which speech should be regulated for the common good. On the one hand, restricting speech in certain contexts can provide key benefits, such as protecting minorities from harm and preventing extremist organizations from recruiting and spreading dangerous hate speech and misinformation. On the other hand, freedom of speech is a fundamental right of individuals to express their opinions and present meaningful political and social discourse.

Facebook has been a key battleground in this debate. As Mark Zuckerberg wrote in a statement detailing the corporation's intended framework for content governance, "One of the most painful lessons I've learned is that when you connect two billion people, you will see all the beauty and ugliness of humanity." Indeed, since nearly its founding day, the company has struggled with the degree to which it bears responsibility for the content that its users post on the platform, including but not limited to, the glorification of violence, incitement to terrorism, and false and misleading political content. Facebook's failures to successfully identify and remove harmful content have in the past enabled grave human rights abuses and acts of violence. In 2017, for example, the Burmese military and other segments of society in Myanmar used Facebook as a platform to spread false information and incite violence against the country's Rohingya minority, which eventually led to the ethnic cleansing of over 1 million people.

Governments, human rights observers, and even private corporations have since criticized Facebook's role in content moderation, going as far as, in some cases, to accuse Facebook of abusing its own users. If societies are ever to collectively decide upon meaningful red lines regarding free speech, they argue, leaving those decisions solely in the hands of corporate executives seems a dubious path to protect the interests of society at large. Facebook has responded to public pressure on this issue, to a limited degree. For example, Facebook has begun releasing transparency reports on its content moderation decisions and has instituted a complex appeals process by which users can dispute determinations on their posts.

Then, in 2020, the company officially launched the Facebook Oversight Board (hence FOB, or the Board), a trust-based body composed of 40 members and tasked with passing final, binding rulings upon Facebook's content moderation decisions. If Facebook takes down a piece of content, and the user who posted that content has exhausted all other mechanisms of appeal, they can then appeal to the FOB as a "supreme court," so to speak, on takedowns. The FOB does not rule upon all cases brought to its attention; rather it selects individual cases which it considers highly emblematic, with the understanding that Facebook will attempt to incorporate its rulings into other, lower-level decisions in the future (although the extent and process by which Facebook will do so is unclear). The Board is also a very recent development; after Facebook announced in 2018 that the FOB would be created, the process of building the institution and selecting members from around the globe was hampered by processual delays, leading to its official initiation in 2020, and its first rulings in January 2021. The Board continues to conduct operations, and at the time of publication had made 10 decisions, the most widely discussed being its recent decision to temporarily uphold Facebook's ban on Donald Trump's account after it was banned for the former president's

role in inciting violence in the wake of the Jan. 6 attack on the U.S. Capitol. The impact and implications of the FOB is discussed in greater detail later in this report.

As more instances of Facebook being used to connect and organize extremists have been exposed, Facebook has faced greater scrutiny regarding its moderation systems. This increasing criticism is hardly unwarranted — despite not being anything remotely resembling a state, the global corporation appears to rival one in its impact upon social and political life. The Myanmar catastrophe in particular exposed the darkest side of this impact, and while Facebook ultimately admitted its role in the genocide of the Rohingya, the implications of its policies and impacts elsewhere in the world must now be addressed such that future episodes of mass violence can hopefully be prevented or mitigated.

One space that has been under-addressed is the role of Facebook's policies in the Syrian conflict. While Facebook was once upheld as a key platform for free, uncensored speech at the beginning of the Arab Spring uprisings, its role in Syria — particularly its impact on Syrian journalists, civilians, and civil society — has become more troubling over time. The rest of this article aims to take stock of Facebook's impact on the conflict in Syria, and to investigate the potential role to be played by the FOB. While the Board ultimately has very little power to force or motivate Facebook to change its policies on the Syrian conflict (or other contexts), it still presents narrow yet important avenues of opportunity for improvement.

The Syrian Regime's Digital Upper Hand

Facebook has long claimed that it strictly maintains no relationships with governments, to maintain the neutrality of its services. And while it is true that it has not to any public knowledge ever made any agreements with figures within or close to the Syrian regime,[1] its existing systems and content moderation policies have given Syria's government a digital upper hand over opposition groups.

Facebook was not designed with the expectation that it would play a role in violent conflicts around the world. Yet for good or ill, that is how it has been used. For example, a number of Syrian-led organizations, both in Syria and around the world, use or have used Facebook to post documentation of human rights abuses so that such information can be stored permanently in a publicly accessible forum. In doing so, the content that they post — at times depicting graphic imagery, such as mutilated bodies or physical attacks — triggers the algorithms designed to find and remove content that violates the company's community standards. And when Facebook removes those posts or the accounts that produced them, all the documentation that has been gathered there simply vanishes with them. This has led to troves of evidence of human rights abuses being tossed into Facebook's algorithmic incinerator, thus preventing their future use in accountability mechanisms. Organizations that have faced takedowns complain that when they have attempted to appeal those decisions, too often there was little to no productive response from Facebook.[2]

However, opposition figures and human rights activists are not Facebook's only users in Syria. The Syrian regime is highly active on social media. Dima Samaro, a MENA-focused expert with an NGO that advocates for digital rights, told MEI, "In Syria, the government mass monitors Syrian citizens on social media, meaning that people can be persecuted, detained, and tortured [for what they post online]. Even people living in exile are still threatened and at risk." Indeed, as early as 2012, the Syrian regime was purchasing tools to monitor dissenters and activists across a variety of platforms, including going as far as infecting protesters' devices with malware.

Facebook's efforts to prevent actors in conflict zones from monitoring the social media usage of activists appear to have had limited impact thus far. At the beginning of the Arab Spring uprisings, it launched the Trusted Partners program, which consists of a network of activists throughout the MENA region who were given special powers to bypass Facebook's content moderation algorithms and refer issues directly to human moderators. The program's purpose was largely two-fold: first, to give Trusted Partners the ability to flag the Facebook accounts of activists and protestors who were detained by armed actors so that their accounts could be blocked before their captors could read potentially incriminating content, and then later to give Trusted Partners the ability to flag content that incites violence or promotes hate. However, while the program was largely successful in locking accounts of detained activists in the MENA region, it failed to significantly improve content moderation. Funding was reportedly weak and poorly structured, Trusted Partners struggled regularly to get a response when flagging content that violated community standards, and Facebook's algorithms even continued to take down Trusted Partners' posts. In the absence of a more effective program to prevent conflict actors from wielding Facebook for violence, the Syrian regime and other actors continue to use Facebook as a tool of surveillance, forcing many Syrians to self-censor at best or face the risk of torture or death for posting their opinions at worst.

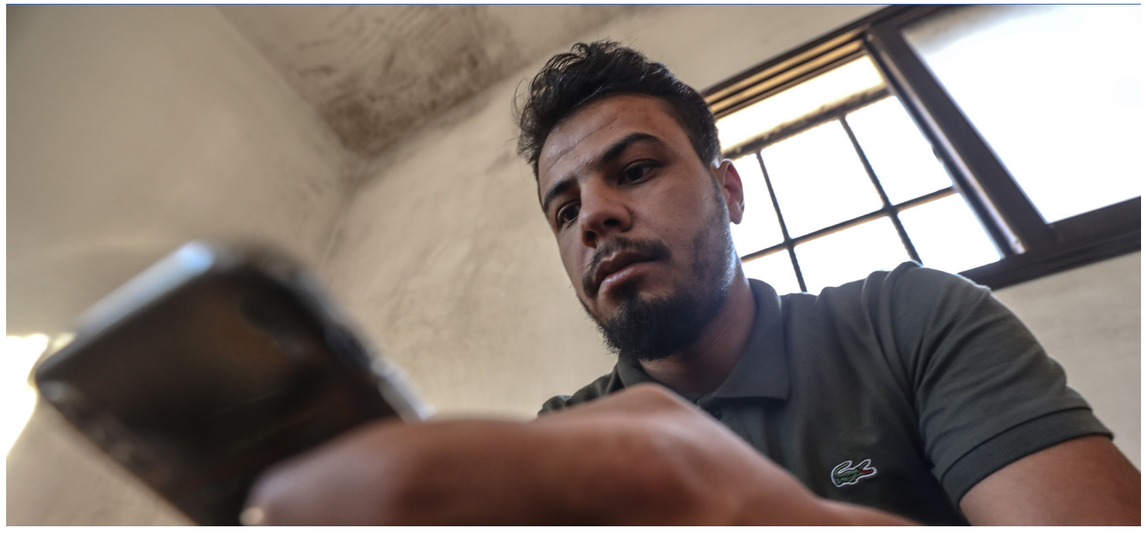

Photo above: Syrian Firas Mansur checks his phone in Idlib, Syria on August 18, 2020. Mansur carries out humanitarian works on civilian victims of the civil war via social media. Photo by Muhammed Abdullah/Anadolu Agency via Getty Images.

At times, the Syrian regime and its supporters have also used Facebook to silence critics more directly. Much has been said in the past decade about the Syrian Electronic Army (SEA), a shadowy legion of pro-Syrian regime hacktivists best known for hacking the Twitter account of the Associated Press and tweeting that the White House had been attacked and President Obama was injured — an overture that caused a $136 billion drop in global stock markets in 2013. But as global media outlets have focused extensively on the SEA's international digital exploits, too little has been said of the group's role in aiding and abetting human rights violations in Syria. During the Arab Spring protests and afterwards, the SEA participated in the monitoring of opposition groups and shared information with the regime such as activists' identities and meeting locations, much of which was obtained via social media and then used to incriminate protesters. Whether the SEA had an explicit connection to the Syrian regime at that time remains unclear, but their efforts were seen in a positive light by government officials, earning the moniker "a real army in virtual reality" from Bashar al-Assad. After largely disappearing from the internet around 2016, the SEA later reemerged with a more explicit connection to the Syrian government. The SEA reportedly has an extensive history of disseminating pro-Assad content on social media and organizing coordinated reports of content posted on Facebook by opposition actors, which results in automated takedowns of said content and endangers opposition social media presence.

Facebook has since taken down many of the SEA's dedicated pages on its platform, but this is unlikely to stop the SEA from monitoring opposition Facebook content and attacking their outlets. Moreover, the author of this article sought out SEA content to include here as examples, only to find that such content has been wiped clean from Facebook's available content archive. While it is for the best that this content is no longer in circulation, the fact that it has been fully removed means that it has also been redacted from the historical record, thus inhibiting potential efforts to bring to justice members and leaders of the SEA. Facebook's disinterest in preserving such content presents a major evidentiary challenge to whatever prosecution efforts may materialize in later years.

Overall, Facebook has still done far too little to prevent forces on the ground from using the platform as a tool to damage their opponents and monitor activists and opposition groups. This leads to a free-for-all environment in which little oversight at the company level allows for abuse of Facebook's tools by actors on the ground, with the Syrian regime and its supporters clearly being the most successful and adept. Finally, the company's propensity to take down content that goes against community standards, without any system to preserve it for future use, is a serious blow to the availability of evidence for justice mechanisms and advocacy efforts.

Photo above: Syrian kids are seen warming around a fire on March 06, 2020 in Idlib, Syria. Civilians have fled due to the ongoing attacks carried out by Assad regime, Russia, and Iranian-backed groups. Photo by Muhammed Said/Anadolu Agency via Getty Images.

Assessing the Role of the Facebook Oversight Board

Since the announcement of its creation in 2018, the FOB has been lauded by some as a real effort toward expert-based, independent content moderation that accounts for both Facebook's internal policies and international human rights norms. But it has been criticized by others for its limited conception of hate speech, its reliance on Facebook for funding and other independence issues, and its relatively limited scope. But regardless of this controversy, the truth is that the FOB simply is not as powerful as it is often described, and moreover its power to rein in violence and hate speech and protect dissent and free speech is limited.

Syria's conflict is a perfect case study into the FOB's limitations. As its website notes, "The board's decisions to uphold or reverse Facebook's content decisions will be binding, meaning Facebook will have to implement them, unless doing so could violate the law." A key clause that could be easily overlooked is the latter portion, which refers to the supremacy of local laws in content decisions. In Syria, the regime has an extensive record of overseeing public communications to further its own ends, including going as far as attempting to block text messages containing words like "demonstration" and "revolution" in 2012. The regime's propensity to directly oversee interpersonal communications endows the state with a legitimate avenue to legislatively check the Board's powers. If, for example, the FOB was to instruct Facebook to restore a Syrian user's post that criticizes the Syrian regime, Facebook could potentially be forced not to comply based on current or future Syrian government legislation. The FOB has yet to review any Syrian users' posts, but the predominance of local legislation outlined in the Board's charter remains a serious check on its power to defend the free speech rights of Syrian dissidents and journalists in particular.

However, arguably the largest check on the FOB's powers is that it can only pass judgement upon content that Facebook has previously chosen to take down. The Board has no power to impact or judge content it has otherwise left standing. And while the Board's charter does indicate that Facebook intends to take the Board's previous decisions into consideration when crafting policy, the charter does not grant the Board any powers to obligate Facebook to implement permanent policy changes in any setting. Thus, if it is within Facebook's policy to leave up content that somehow contributes to hate or take down content that should be left up for purposes of documentation, the FOB cannot obligate Facebook to the contrary. The FOB also seems to have no direct control over Facebook's algorithmic moderation systems, which are the platform's crucial first line of defense in content moderation.

Syria provides a demonstrative case study. While Facebook claims that its algorithms have been well-trained to halt the flow of terrorist content, questions remain about the proliferation of hate speech. Facebook's community standards list 10 protected characteristics used to moderate hate speech: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity, and serious disease. Not listed among these is conflict affiliation and profession. While it is crucial to allow criticism of political and professional groups, there is a line to draw: for example, these community standards seem to indicate that someone posting a phrase such as "All who oppose Assad are insects" would be permitted, even though this phrase indicates inferiority of Syria's political opposition and could contribute to real-life violence. Professions are protected, but only when they are paired with another protected characteristic, meaning the blanket statement "All journalists are dogs" would also seemingly not be taken down. Finally, numerous people interviewed for this paper argued that Facebook has underinvested in Arabic language content moderation, and that existing algorithms and human staff lack the capacity to detect and remove the broad cross-section of hate speech which proliferates in Arabic's many regionally spoken dialects and vernaculars. As a result, even grave and clear violations of Facebook's community standards often continue to circulate among Arabic-speaking communities on the platform.

The bottom line here is that there are forms of hate speech that are permitted by Facebook's policies or that slip around algorithmic blockers, including speech that could have a real impact in conflict. And it is precisely because this content is not taken down under Facebook's existing policies that the FOB will never be able to rule upon it. Unless Facebook's content moderation systems change such that this content is taken down, it will never be subjected to independent review by the FOB.

As a result, the FOB's real powers to transform and improve Facebook's impact upon Syria is quite limited. Without the power to review a wider and less selective range of content, the Board is limited to only a small sub-section of the content that Facebook struggles to moderate and cannot change the broader schema of speech that proliferates on the platform. These conditions also largely prevent the FOB from impacting the free-for-all environment that exists on Facebook in Syria. As a result, different parties to the conflict, most notably the Syrian regime, will continue to use Facebook's tools in pursuit of their own goals and ambitions, including in the service of violent intentions, such as using the social media content posted by critics and activists to incriminate them.

At the end of the day, the Board was created by Facebook to limit public and government concerns about Facebook's content moderation systems rather than to promote substantive change, and much more must be done to improve Facebook's impact upon Syria's conflict and other zones of frequent violence around the world. Nevertheless, the FOB still presents narrow opportunities to bring about small, though productive changes. Going forward, members of the Board should implement the following targeted action recommendations when going about their work:

Recommendations for the Facebook Oversight Board:

- While Facebook has not yet contributed to a singular event in Syria as extreme as the Rohingya crisis, the platform is still used frequently in Syria and elsewhere as a tool to abet and implement severe human rights abuses. Therefore, the Board should step into its capacity as a human rights-focused oversight organ and engage closely with Facebook to implement new content moderation systems that prevent Facebook from being used to harm people. It should advocate wherever possible for a human rights-centric framework for content governance that prioritizes protecting civilians, respects the rights of victims, and terminates the use of Facebook as a tool for surveillance and incrimination. The Board should also notify the public in clear and certain terms when Facebook fails to live up to its responsibilities to protect the rights and interests of its users in Syria and other conflict zones.

- When ruling upon hate speech posts written in less widely spoken languages, the FOB should recommend that Facebook earmark greater resources toward the aggregation and

incorporation of data sets on hate speech in these languages. It should include — as is often the case in Arabic — local dialects, as speakers of local vernaculars are often members of communities that are most disproportionately impacted by Facebook's language divide in content moderation. Board members should stress that Facebook's lack of Burmese moderation controls played an important role in its complicity in the attempted genocide of the Rohingya, and that greater language investment, including in less widely spoken languages, is an important mechanism to prevent similar tragedies.

- If given the opportunity (either from a specific relevant case or by request from Facebook), the FOB should engage directly with Facebook on the issue of documentation of human rights abuses being removed from the platform. The FOB should use any available means to encourage Facebook to archive content that depicts human rights abuses or other activities in contravention of international law so that it can later be shared with the appropriate authorities. The Board should note to Facebook that its platform can have a real, positive impact on accountability mechanisms if it shows greater willingness to collaborate with governments on the sharing of criminal evidence in cases which concern grave human rights abuses.

- The Board should encourage Facebook to implement wider and more effective democratization processes when establishing its policies on content moderation. In particular, Facebook should include its Trusted Partners in decision-making in a robust and systematic way, rather than relying on them only to report content. The inclusion of local perspectives can only increase the company's efficacy when removing dangerous or inciteful content. Finally, if an opportunity arises within the constraints of the FOB's charter, the Board should recommend that Facebook expand the Trusted Partners program, offer Trusted Partners greater and more well-structured funding, and improve the speed and efficiency of the mechanisms by which Trusted Partners can report illegal or harmful content.

- In 2018, Facebook released a Blueprint for content governance and enforcement that aims to respond proactively to the fact that salacious (and often offensive) content typically receives the most engagement from users, which feeds cycles of information violence. The company declared in the Blueprint an intention to "penalize" content that approaches the line of violating Facebook's policies so that such content receives less engagement rather than more. This is an important and promising step toward the limitation of hate speech. However, there seems to have been little actionable progress regarding the implementation of the Blueprint. The FOB should request that Facebook provide the Board with periodic, timely, goal-oriented reports on its implementation.

Acknowledgements

The author graciously thanks the following experts for contributing their perspectives in the creation of this piece: Dima Samaro, Middle East and North Africa policy analyst at Access Now; Emerson Brooking, resident senior fellow at the Atlantic Council's DFRLab; Gabriel Young, Ph.D. candidate in History and Middle Eastern Studies at New York University; Mohammad Al-Abdallah, Executive Director of the Syria Justice and Accountability Centre; and Noura Al-Jizawi, Researcher at Citizen Lab, University of Toronto.

About the author

Leo Hochberg is Graduate Research Fellow with the MEI Cyber Program and a Masters student in Conflict Studies and Human Rights at Utrecht University. His research focuses include refugees and human (in)security, transitional justice, and MENA regional cyber threats. The views expressed in this piece are his own.

Endnotes

1. Facebook has no public relationship with the Syrian government or its close affiliates; however, Facebook's internal decision-making processes operate behind a high wall of secrecy. Thus, while the author of this report and the general public are unaware of any existing agreements, relationships, or biases in decision-making in this regard, that does not necessarily mean they do not exist — we simply do not know whether Facebook or any of its key employees are predisposed to support the Syrian regime, or other violent actors in Syria's conflict.

2. Access Now, a digital rights NGO, operates a help line for journalists, news outlets, and civil society organizations to report the closure of their accounts and seek redress. While it forwards these requests to Facebook and advocates for reinstating accounts, Facebook does not respond to all requests and many of the accounts are never reopened. Information provided by interview with Dima Samaro, a MENA digital rights expert at Access Now.