Considering the Whole Ecosystem in Regulating Terrorist Content and Hate Online

https://www.e-ir.info/2019/09/18/considering-the-whole-ecosystem-in-regulating-terrorist-content-and-hate-online/

AMY-LOUISE WATKIN, SEP 18 2019

The last few years have seen a range of proposals to counter terrorist and extremist content online. From the European Commission's Code of Conduct on Countering Illegal Hate Speech Online in 2016 and its 2018 proposed Regulation for preventing the dissemination of terrorist content online, to the introduction of the German Network Enforcement Act (NetzDG) in 2017, with a similar French law likely to be implemented soon, and the UK's Online Harms White Paper in 2019, a range of strategies have been advanced. One recurrent idea is a requirement for faster removals, supported by large fines in the event of non-compliance, and the UK's Online Harms White Paper proposed the creation of a new independent regulatory body.

This article is going to argue that there are two crucial considerations that require greater thought when putting forth proposals. The first is the enormous variation that exists between the social media platforms, file-sharing sites and instant messaging services that are being used, in terms of their capabilities and resources, as well as their motivations to comply. The second is the large variety of services these organisations supply and the finding that the platforms and sites are interconnected yet not used homogeneously. The concluding argument will be that the whole ecosystem requires consideration in future proposals.

The variety in size and resources

The first significant difference, which has an important impact on an organisation's ability to counter terrorist and extremist exploitation of their sites, is its number of staff and revenue. For example, Facebook has 2.2 billion users, with almost 40,000 full-time employees. Thirty thousand of these work in safety and security with half employed as content reviewers globally. Although also considered a major platform, Twitter is a lot smaller with 126 million active daily users and 3,900 global employees, 1500 of whom work on policy enforcement of content moderation. Twitter, however, is enormous compared to many file-sharing sites and alternative platforms whose staff are often based in just one country. JustPaste.it, a file-sharing site that claims to have 5 million users, has been known to be managed solely by its creator Mariusz Zurawek. Gab, an alternative social media platform, is thought to have only a handful of employees and approximately 1 million users. Gab is an example of a platform that does not yet fall under Germany's NetzDG law, since this only applies to organisations with 2 million registered users in Germany. The major social media platforms bring in large revenues from running advertisements on their sites, for example, in January 2019, Facebook reported ad revenue of $16.6 billion. JustPaste.it and Gab on the other hand, rely on the less stable strategies of selling premium accounts, and Gab also sell merchandise and accept donations. Therefore, some organisations have large global teams of employees ranging from safety and security experts to content reviewers, while others have very few, if any employees working on this specifically, and some organisations are dependent on keeping advertisers on their platform for revenue, with others reliant on their users for revenue. With this variation between organisations, it is likely that regulatory strategies will have to differ depending on the size and resources of the platform in question.

There are also significant differences in the organisations' missions and values. Facebook, Twitter and YouTube have stated consistently that they work to balance freedom of expression with the safety of their users. Facebook in particular argue that they will not retain their user-base and advertisers if their platform is not safe. They also state

that whilst they have their own global policies – they follow local laws to prevent being blocked by authoritarian governments that would remove their services from whole populations who rely on them. The aim of JustPaste.it, on the other hand, is to provide "the quickest way to share text & images with other people" whilst offering a "high level of privacy for both writers and readers". Finally, Gab was created with the aim of running a site powered by its users, with an emphasis on providing user privacy, and prides itself on its lack of censorship. Gab recently became a decentralised platform which makes it more resistant to content moderation efforts. While the major platforms assert that, in order for everyone to safely have a voice, some speech (particularly that aimed at protected groups) has to be regulated, alternative platforms such as Gab take the view that speech is not truly free where regulation exists.

Variation can also be seen across the policies these organisations have in place. Facebook, Twitter and YouTube have already created and implemented technology that they claim proactively removes large volumes of content that violates their policies, despite members of the UK Home Affairs Committee calling out all three platforms for still containing a significant portion of violating content on their sites in April 2019. Facebook argue, however, that the number of takedowns is not the best metric to go by, rather they believe it is better to focus on takedowns of content that are likely to receive the most views. JustPaste.it has not implemented the same proactive technology and is limited by their lack of staff, however, it works with Tech Against Terrorism and the Global Internet Forum to Counter Terrorism (GIFCT) (for example, on the hashing database) to try to counter terrorist content on its site. Gab's Terms of Service reveal a reactive approach as they "do not review material before it is posted on the website and cannot ensure prompt removal of unlawful material after it has been posted". These examples reveal the current existing spectrum of organisations with strategies that range from proactive to reactive, with sites like JustPaste.it falling somewhere in between.

The variety of services

The other crucial consideration is how terrorist and extremist groups are utilising these services. With the major platforms getting progressively better at suspending terrorist accounts (Conway et al. 2017; Berger & Perez 2016), and a recent crackdown on those associated with the far-right, a range of other platforms and file-sharing sites are essential to the efforts of these groups. Research by Weirman and Alexander revealed that despite Twitter getting better at removing violating content posted by the so-called Islamic State (IS), the group has adapted (2018). They began to use Twitter, instead, to share news sources, particularly those that validate the group's stance and in isolation do not necessarily violate any policies. This research along with other studies (Macdonald et al. 2019; Conway et al. 2017) have shown that IS also use Twitter to redirect their followers to file-sharing sites. Shehabat and Mitew (2018) studied some of these file-sharing sites and found that IS post large quantities of content there that would cause removal and suspension on the major platforms. They said that these file sharing sites allow "for fluid and anonymous information aggregation, curation, and dispersal" (*Ibid*., p.97*)*. Further to this, research by Clifford and Powell (2019) found that IS have also used major platforms to direct their followers to the online instant messaging service Telegram which prides itself on user privacy, and allows access to both public channels and private chat groups. Here they use it to "communicate with like-minded supporters across the world, disseminate official and unofficial IS media, and provide instructional material for operations" (*Ibid.,* p.3). All this research highlights two key points. The first is that these sites are not being used homogeneously –some are used to signpost, others as repositories or for communication. The second is the interconnectedness of the groups' strategies, using the different platforms to signpost to other sites. Effective regulation will therefore have to tackle the file-sharing sites in order to remove the repository of materials that are posted there but will also have to tackle the major platforms that are currently being used to disseminate the URLs signposting to the file-sharing sites. Research by Alexander and Braniff (2018), however, identifies some of the downfalls of focusing solely on content removal and suggests a range of strategies that can be used in addition to try to marginalize the efforts of these groups depending on how they are exploiting specific platforms.

Due to the major platforms' recent crackdown on far-right groups, there is a similar emerging pattern by these groups of migration to Telegram and alternative platforms to communicate and disseminate content. One example is the UK-based far-right group Britain First, which moved to Gab when they were removed from Facebook and Twitter. Recent research studied whether the move to the alternative platform led to any changes in their follower count and content dissemination strategies. It found that although the group struggled to regain the enormous following that they had on

the major platforms, some of the content posted by the group was becoming more extreme than it had been on Facebook (Nouri et al., 2019), possibly a result of the platform's stance on free speech and censorship. Therefore, regulatory strategies also have to be weary of the whack-a-mole effect that comes with focusing on the major platforms and the consequences it has on content being disseminated elsewhere.

There is, therefore, significant variation between both the capabilities and resources, as well as the motivations of these organisations to regulate terrorist and extremist content. This was demonstrated in this article with only a small handful of platforms and sites. An analysis by Tech Against Terrorism found evidence of terrorist groups using more than 330 platforms with 25 of the top 50 most-used platforms by IS being small or micro-platforms, suggesting that the actual variation between platforms is likely to be much bigger. A one size-fits-all regulatory approach is, therefore, likely to result in regulatory strategies working well for some organisations but creating enormous struggles and vulnerabilities for others. File-sharing sites, social media platforms and instant messaging services are all being utilised simultaneously for different purposes, and are interconnected in use through the posting of URL links signposting to one another. The groups are learning that they can avoid removal from the major platforms by circulating non-violent content supporting their cause (that research has found can be just as persuasive as violent content), whilst using the less staffed and resourced file-sharing sites to post large quantities of content that would violate the major platforms policies.

Conclusion

Regulatory strategies are going to have to focus on more than just the major platforms, the removal of content that is typically considered 'dangerous' and 'inciting'[1], and fast removals that require a lot of staff and proactive technology. A strategy that focuses exclusively on the social media giants will prove ineffective because they only make up one small part of the many platforms, sites and messenger services that constitute the whole ecosystem of organisations that are currently in use by terrorist and extremist groups. A regulatory strategy that only focuses on the removal of inciteful content will overlook the non-inciteful content that is able to remain and provide a safe space for the groups to post URL links to inciteful content elsewhere. It could also lead to the whack-a-mole effect with groups migrating to parts of the internet that are less censored or more difficult for law enforcement to monitor. A regulatory strategy that focuses exclusively on the removal of content within a short timeframe will prove ineffective because of the enormous volume of content these organisations have to sort through, possibly risking a 'better safe than sorry' attitude to content removal. Finally, a regulatory strategy that focuses exclusively on the imposition of fines in the event of non-compliance risks the major companies treating them as just another business cost. It may also be difficult to enforce fines on organisations that are registered outside of the jurisdiction, while smaller companies that are willing to comply but face barriers when it comes to capabilities and resources may be punished.

Focusing on the whole ecosystem is going to require regulatory frameworks to include all of the platforms, file-sharing sites and instant messenger services, as opposed to just the major platforms or platforms with a specific number of users, with consideration of all the specific ways that they are each exploited (e.g., as a repository, signposting, communication etc.). It should ensure that companies with large volumes of terrorist or extremist content are not able to slip through the cracks by having a smaller user-base. The regulation will also have to take into account that some organisations are more equipped and/or motivated to comply than others. The UK's Online Harms White Paper proposed a regulator to provide advice and guidance, and to facilitate the sharing of technological tools and best practice, which would be helpful for sites such as JustPaste.it. The White Paper also proposed the disruption of business activities and ISP blocking. Although this raises serious concerns –for example, its effect on freedom of speech, particularly on platforms where there is a significant portion of lawful content—it is unlikely that other strategies will have any effect on uncooperative organisations, particularly those that are outside of the jurisdiction, and therefore may be required as a means to incentivize these organisations or as a last resort in extreme cases. Finally, the regulations need to consider that the number of takedowns, speed of takedowns, and removal of content that falls under very specific policy terms such as 'incitement' is not necessarily the best form of action. It should investigate other types of content, additional strategies to removal and other removal strategies, such as the one Facebook put forward of focusing on removing content that has the potential to be widely viewed.

Notes

[1] For example, see Facebook Community Standards on Violence and Criminal Behaviour.

Bibliography

Alexander, A., and Braniff, W. (2018) Marginalizing violent extremism online. *Lawfare Blog.* Accessed 9 September 2019 via https://www.lawfareblog.com/marginalizing-violent-extremism-online

Berger, J. M., & Perez, H. (2016) *The Islamic State's Diminishing Returns on Twitter: How Suspensions are Limiting the Social Networks of English-speaking ISIS Supporters*. George Washington University.

Bickert, M. (2017) Facebook's Community Standards: How and Where We Draw the Line. *Facebook*. Accessed September 2017 via https://newsroom.fb.com/news/2017/05/facebooks-community-standards-how-and-where-we-draw-the-line/

Bickert, M., and Fishman, B. (2018) Hard questions: What are we doing to stay ahead of terrorists?*Facebook Newsroom.* Accessed November 2018 via https://newsroom.fb.com/news/2018/11/staying-ahead-of-terrorists/

Clifford, B., and Powell, H. (2019) Encrypted Extremism: Inside the English-speaking Islamic State ecosystem on Telegram. *Program on Extremism, The George Washington University.*

Conway, M., Khawaja, M., Lakhani, S., Reffin, J., Robertson, A., & Weir, D. (2017) Disrupting Daesh: measuring takedown of online terrorist material and its impacts. VoxPol Network of Excellence for Research in Violent Online Political Extremism.

Facebook (2018) Hard Questions: The Line Between Hate and Debate. *Facebook Newsroom.* Accessed August 2018 via https://newsroom.fb.com/news/2018/08/the-line-between-hate-and-debate/

Facebook (2019a) A conversation with Mark Zuckerberg, Jenny Martinez and Noah Feldman. *Facebook Newsroom.* Accessed July 2019 via https://newsroom.fb.com/news/2019/06/mark-challenge-jenny-martinez-noah-feldman/

Facebook (2019b), Community Standards, Violence and Criminal Behaviour. Accessed 16 September 2019 via https://www.facebook.com/communitystandards/violence_criminal_behavior

Gilbert, D. (2019) Here's how big far right social network Gab has actually gotten. *Vice*. Accessed 28 August 2019 via https://www.vice.com/en_us/article/pa7dwg/heres-how-big-far-right-social-network-gab-has-actually-gotten

Lee, D. (2018) Tech firms hail 'progress' on blocking terror. *BBC*. Accessed 30 August 2019 via https://www.bbc.co.uk/news/technology-44408463

Macdonald, S., Grinnell, D., Kinzel, A., and Lorenzo-Dus, N. (2019) A study of outlinks contained in tweets mentioning 'Rumiyah'. *The Global Research Network on Terrorism and Technology*

Minshall, K. (2019) UK Home Affairs Committee: Hate crime and its violent consequences. *House of Commons.* Accessed 30 August 2019 via http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/home-affairs-committee/hate-crime-and-its-violent-consequences/oral/100660.pdf

Nouri, L., Lorenzo-Dus, N., and Watkin, A. (2019) Following the whack-a-mole: Britain First's Visual Strategy from Facebook to Gab. *The Global Research Network on Terrorism and Technology*

Potts, N. (2019) UK Home Affairs Committee: Hate crime and its violent consequences. *House of Commons.* Accessed 30 August 2019 via http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/home-affairs-committee/hate-crime-and-its-violent-consequences/oral/100660.pdf

Robins-Early, N. (2019) Like ISIS before them, far-right extremists are migrating to Telegram. *Huffington Post.* Accessed 28 August 2019 via https://www.huffpost.com/entry/telegram-far-right-isis-extremists-infowars_n_5cd59888e4b0705e47db36ef

Rosen, G. (2018) How are we doing at enforcing our community standards? *Facebook Newsroom.* Accessed August 2019 via https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/

Shaban, H. (2019) Twitter reveals its daily active user numbers for the first time. *The Washington Post*. Accessed 30 August 2019 via https://www.washingtonpost.com/technology/2019/02/07/twitter-reveals-its-daily-active-user-numbers-first-time/?noredirect=on&utm_term=.e92a96985f22

Shehabat, A., & Mitew, T. (2018) Black-boxing the black flag: anonymous sharing platforms and ISIS content distribution tactics. *Perspectives on Terrorism*, *12*(1).

Tech Against Terrorism (2017) Tech Against Terrorism at Chatham House. *Tech Against Terrorism*. Accessed October 2017 via https://techagainstterrorism.org/2017/07/12/tat-at-chatham-house/

Tech Against Terrorism (2019a) Analysis: The use of open-source software by terrorists and violent extremists. Accessed 3 September 2019 via https://www.techagainstterrorism.org/2019/09/02/analysis-the-use-of-open-source-software-by-terrorists-and-violent-extremists/

Tech Against Terrorism (2019b) ISIS use of smaller platforms and the DWeb to share terrorist content.*Tech Against Terrorism*. Accessed 30 August 2019 via https://www.techagainstterrorism.org/2019/04/29/analysis-isis-use-of-smaller-platforms-and-the-dweb-to-share-terrorist-content-april-2019/

Timberg, C., Harwell, D., Dwoskin, E., and Brown, E. (2018) From Silicon Valley elite to social media hate: The radicalization that led to Gab. *The Washington Post.* Accessed July 2019 via https://www.mercurynews.com/2018/11/01/from-silicon-valley-elite-to-social-media-hate-the-radicalization-that-led-to-gab/

Waters, G. & Postings, R. (2018) Spiders of the Caliphate: Mapping the Islamic State's global support network on Facebook. *Counter Extremism Project.* Accessed June 2018 via https://www.counterextremism.com/sites/default/files/Spiders%20of%20the%20Caliphate%20%28May%202018%29.pdf

Weirman, S., & Alexander, A. (2018) Hyperlinked Sympathizers: URLs and the Islamic State.*Studies in Conflict & Terrorism*, 1-19.

About the author:

Amy-Louise Watkin is a PhD candidate at Swansea University and a member of the Cyber Threats Research Centre.