Routing the Extreme Right: Challenges for Social Media Platforms

Maura Conway

Abstract

Between 2014 and 2017, the Islamic State maintained vibrant communities on a range of social media platforms. Due to aggressive account and content takedown policies by the major platforms, these visible communities are now almost non-existent. Following the March 2019 Christchurch attack, the question as to why major platforms cannot rout the extreme right in the same way has repeatedly arisen. In this article, *Maura Conway* explores why this is not as straightforward as it may seem.

Introduction

Globally, state actors came to swift agreement on the terrorist status of the Islamic State (IS)[1] and, after initial hesitance from major social media companies,[2] they concurred that terrorist content was not welcome on their platforms.[3] This sort of agreement was possible not just because of IS's terrorist activity, but also its group structure. At its height, IS was an organisation with a defined leadership structure, a bureaucracy that included a complex media production and circulation system, which coalesced around a set of core ideological commitments, and a terrorism strategy that was part of what it viewed as a broader state-building and expansion strategy. The extreme right, on the other hand, might be termed a 'scene', 'milieu' or 'ecology', composed of a fast-changing and complex overlapping of individuals, groups, movements, political parties and media organs – both online and offline – espousing extreme nationalist, National Socialist, fascist, white supremacist and/or 'alt-right' ideology. This fast-changing 'scene' is not dominated by one or even a small number of groups, as is the case with violent jihadism.

---

[1] UN Security Council Resolution 2253, 17 December 2015, S/RES/2253.

[2] Jessica Stern and J M Berger, *ISIS: The State of Terror* (London: HarperCollins, 2015), pp. 138–39; YouTube, 'Dialogue with Sen. Lieberman on Terrorism Videos', blog post, 19 May 2008, http://googlepublicpolicy.blogspot.ie/2008/05/dialoguewith-sen-lieberman-on.html>, accessed 23 January 2020.

[3] See, for example, this detailed analysis of Twitter's response: Maura Conway et al., 'Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts', *Studies in Conflict and Terrorism* (Vol. 42, No. 1–2, 2018), pp. 141–60. It is worth noting that the Islamic State (IS) and its supporters still make considerable daily efforts to establish new footholds on major platforms.

And even fewer are officially designated as terrorist groups in any jurisdiction (in the UK, for example, just one extreme-right group, National Action, has such a designation).[4]

Once it was accepted by most major international actors, including major Western social media companies, that IS was a terrorist organisation about which something had to be done, it was relatively straightforward to make that happen. This was largely due to its branding strategy. The organisation branded all of its official online content with logos – this, while initially contributing to its online rise, facilitated its eventual downfall. Getting rid of IS content from major social media platforms necessitated the development of tooling to find and remove IS-branded content, and due to the organisation's cohesive branding, this method proved highly effective. In a late 2017 blog post, Facebook announced that it was able to remove 99% of IS and Al-Qa'ida material prior to it being flagged by users due to advances in machine learning and artificial intelligence (AI).[5] While some right-wing extremist content is associated with particular groups and branded as such, the large number of such groups, on the one hand, and the fact that the vast majority of right-wing extremist content is not branded, on the other, complicate this approach from the outset. This, coupled with the important caveat that very few right-wing extremist groups are designated as terrorist organisations, poses a significant challenge. This hindrance is acknowledged in the Facebook blog post, which states that 'we're a long way from being able to rely on machine learning and AI to handle the complexity involved in assessing hate speech' and reaffirms the importance of 'the eyes and ears of everyone on [the] platform' to report potential hate content.[6]

These difficulties are compounded by right-wing extremists' heavy co-optation of meme culture. Memes can take a variety of forms, including catchphrases, easy-to-digest captioned images and quirky animated GIFs. They are widely popular online, with many achieving viral status.[7] Commonplace, for example, are images overlaid with 'humorous' text. James Alex Fields Jr, sentenced to life imprisonment in March 2019 for killing Heather Heyer by driving his car into a crowd of counter protesters at the 'Unite the Right' rally in Charlottesville, shared a meme on Instagram months prior to the rally. It showed a car driving through a crowd of people described as protesters, and the post was captioned: 'You have the right to protest, but I'm late for work'.[8]

---

[4] National Action was proscribed as a terrorist organisation in the UK in December 2016, with its alternative names 'Scottish Dawn' and 'NS131 (National Socialist Anti-Capitalist Action)' added in September 2017. See UK Home Office, *Proscribed Terrorist Organisations* (London: Home Office, 2019), pp. 18–19. In May 2019, Canada followed up by listing neo-Nazi group Blood & Honour and its armed branch, Combat 18, as terrorist organisations. See Public Safety Canada, 'Currently Listed Entities', <https://www.publicsafety.gc.ca/cnt/ntnl-scrt/cntr-trrrsm/lstd-ntts/crrnt-lstd-ntts-en.aspx#60>, accessed 20 January 2020.

[5] Monika Bickert and Brian Fishman, 'Hard Questions: Are We Winning the War On Terrorism Online?', Facebook Newsroom, 28 November 2017, <https://newsroom.fb.com/news/2017/11/hard-questions-are-we-winning-the-war-on-terrorismonline/>, accessed 20 January 2020.

[6] *Ibid*.

[7] 4chan has generated many of the Internet's most persistent and pervasive memes, including 'LOLcats' (pictures of cats with humorous overlay text) and 'Rickrolling' (unexpected appearances of the music video for Rick Astley's 1987 hit 'Never Gonna Give You Up').

[8] Commonwealth of Virginia vs. James Alex Fields Jr, 'Commonwealth's Motion *in limine* to Admit Two Instagram Posts', Circuit Court for the City of Charlottesville, 29 November 2018.

Together, the memes, specialised jokes and jargon dominating the online spaces of the extreme right serve as a constant stream of highly distilled ideological thought, reinforcing the extremist beliefs of these virtual communities, while also acting as a means of identity creation and formation for users both new and old.[9] However, taken separately, they are not interpretable as terrorist content. At one point in 2015, more than 50% of IS's online content was even more anodyne than this, consisting of a stream of images of fruit-laden market stalls, well-equipped hospitals, children's playgrounds, and bucolic countryside scenes.[10] On the face of it, this was not interpretable as terrorist content either, but the fact that it was branded by IS meant that it could therefore be removed on the basis of the organisation's terrorist status.[11]

While it is true that IS has been severely disrupted on major social media platforms, this is not true for violent jihadism more generally. Instead, research shows that a process of differential disruption is taking place whereby a variety of other violent jihadi groups – including Hayat Tahrir al-Sham, the Syrian Liberation Front (formerly Ahrar Al-Sham) and the Taliban – are much less impacted.[12] This can be explained by at least three reasons: there is less of an emphasis on removing this content because of the lack of consensus on the extremist and/or terrorist status of some of these groups; they are not viewed as posing the same level of threat as IS, either in the 'real world' or online; and their content is less well known to (and therefore less discoverable by) social media companies and their tooling. Thus, though social media companies' responses to IS activity have been positive, the same cannot be said for their takedown of violent jihadi content more generally. Extending such an analysis to the whole range of online content which either valorises or espouses core violent jihadi commitments would undoubtedly turn up an even greater volume of content on major platforms.

Demands to disrupt the extreme right on major social media platforms are much more akin to this disruption of violent jihadism rather than of IS specifically. Therefore, they are likely to have similarly uneven results.

The Shift in the 'Overton Window' and Monetisation of Extreme-Right Content

There are numerous heads of state, political parties, major media outlets and large voter constituencies throughout the Western world (the same is true elsewhere too) that are supportive to greater and lesser degrees of the extreme right. In particular, the Donald Trump administration in the US is venerated by many on the extreme right, not just in

---

[9] Jordan Eschler and Amanda Menking, '"No Prejudice Here": Examining Social Identity Work in Starter Pack Memes', *Social Media + Society* (Vol. 4, No. 2, 2018), pp. 1–13.

[10] Charlie Winter, *Documenting the Virtual Caliphate* (London: Quilliam, 2015), pp. 30–37.

[11] However, it should be noted that even the takedown of IS-branded non-violent content is not uncontentious. See Europol, 'On the Importance of Taking-Down Non-Violent Terrorist Content', *VOX-Pol*, 8 May 2019, <https://www.voxpol.eu/onthe-importance-of-taking-down-non-violent-terrorist-content/>, accessed 20 January 2020.

[12] *Ibid*. See also Maura Conway et al., 'A Snapshot of the Syrian Jihadi Online Ecology: Differential Disruption, Community Strength, and Preferred Other Platforms', *Studies in Conflict and Terrorism* (forthcoming 2020).

America, but globally.[13] Not only is President Trump seen as endorsing right-wing extremist views,[14] including by right-wing extremists themselves,[15] but he has also had a hand in widely mainstreaming them. The ideas of the Christchurch and El Paso shooters are freely invoked on mainstream social media platforms,[16] as well as in traditional media.[17] For example, in an online post, Patrick Crusius, the alleged El Paso shooter, described the purpose of his attack as warding off the 'Hispanic invasion of Texas', which is a commonplace talking point on the US president's preferred television station, Fox News,[18] and a trope that Trump himself has employed repeatedly.[19] This shift in the 'Overton Window' (the range of ideas tolerated in public discourse) make it difficult for social media companies to respond effectively to right-wing extremist content and activity on their platforms because to do so will increasingly be framed as reflecting a political bias against the right more broadly. This challenge was illustrated by the White House's decision to convene a meeting with Internet companies for a discussion on violent online extremism in the wake of the August 2019 El Paso shooting, while simultaneously circulating drafts of a proposed executive order that would address alleged anti-conservative bias by the same companies.[20] Unsurprisingly, IS was (and is) lacking any equivalent public shows of support by world leaders or their governments.

There is also a suspicion that profit motives play a role in some platforms' attitudes to the availability of extreme-right content on their services. In other words, the shift in the Overton Window presents a profit opportunity. All major social media companies generate the bulk of their revenue from advertising, which tends to be priced on the basis of the amount of interaction their content generates. What this means, in effect, is that the possibility for some platforms to profit from extreme-right activity is considerable. Even at the height of their social media activity, IS content and supporter accounts did not have the kind of traction or reach which is commonplace amongst today's extreme-right influencers. To illustrate this, the average number of followers among IS supporters on Twitter in 2014 – well before significant disruption by Twitter of supporter accounts

[13] J M Berger, *The Alt-Right Twitter Census: Defining and Describing the Audience for Alt-Right Content on Twitter* (Dublin: VOX-Pol, 2018), p. 6.

[14] For example, see then British Prime Minister Theresa May's response to Trump's retweets of Britain First's then Deputy Leader Jayda Fransen in November 2017: *BBC News*, 'Donald Trump Wrong to Share Far-Right Videos – PM', 29 November 2017.

[15] Mary Papenfuss, 'Proud Boys Deem Portland a Success Because Trump Sided with Right-Wing Extremists', *HuffPost*, 18 August 2019.

[16] Jacob Davey and Julia Ebner, *The 'Great Replacement': The Violent Consequences of Mainstreamed Extremism* (London: Institute for Strategic Dialogue, 2019).

[17] For example, see Cristina Lopez, 'Fox News Talked More About Migrant "Invasion" Just Before Election Than in Past 3 Years Total', *HuffPost*, 9 December 2018; Edward Helmore, 'Fox News Host Compares Migrants Entering US to Nazis', *The Guardian*, 15 August 2019.

[18] Jeremy W Peters et al., 'How the El Paso Killer Echoed the Incendiary Words of Conservative Media Stars', *New York Times*, 11 August 2019.

[19] For example, @realDonaldTrump, 'Many Gang Members and some very bad people are mixed into the Caravan heading to our Southern Border. Please go back, you will not be admitted into the United States unless you go through the legal process. This is an invasion of our Country and our Military is waiting for you!', 2:41 pm, 29 October 2018, <https://twitter.com/realDonaldTrump/status/1056919064906469376>, accessed 23 January 2020. See also *Reuters*, 'Trump Calls Migrant Caravans "Invasion" at Campaign Rally', 9 May 2019.

[20] Margaret Harding McGill and Daniel Lippman, 'White House Drafting Executive Order to Tackle Silicon Valley's Alleged Anti-Conservative Bias', *Politico*, 7 August 2019.

kicked in – was 1,004, while the same figure for alt-right accounts on Twitter in 2018 was 6,779. The median followers for IS supporter accounts was 177, and for alt-right accounts it was 561.[21]

YouTube, in particular, has come under fire with respect to attention-based monetisation. In 2019, the *Washington Post* reported on the experiences of YouTube moderators and their claims that abusive posts by popular creators were subject to less stringent moderation than posts by those with less lucrative channels. In June 2019, and only after a public outcry, YouTube demonetised Steven Crowder, an extreme-right commentator with over four million subscribers, for repeated verbal abuse of Carlos Maza, a homosexual Hispanic journalist. The incident, which coincided with an already-planned change in YouTube's hate speech policy,[22] displayed significant inconsistencies in how the platform enforces its rules. At one point, it appeared to interpret its own public policy on hate speech and harassment in four different ways within 24 hours. According to the *Washington Post*, moderators had already escalated their concerns about Crowder's activity on the platform to YouTube at least one month prior. 'The consensus on the floor was that the content was demeaning and it wasn't safe', said one of the moderators interviewed. 'YouTube's stance is that nothing is really an issue until there is a headline about it'. A current moderator was more explicit: 'The picture we get from YouTube is that the company has to make money – so what we think should be crossing a line, to them isn't crossing one'.[23]

The US's Outlier Legal Context and Its Impact

The largest and most influential social media companies – and most of the fringe online platforms hosting the most virulently extreme-right content – are not 'Western', but specifically American. The First Amendment to the US Constitution protects hateful speech, a category into which much right-wing extremist content falls. In addition, the US Supreme Court ruled, in Brandenburg v. Ohio, which overturned the conviction of a Ku Klux Klan member, that the government may not ban 'the mere abstract teaching … of the moral propriety or even moral necessity for a resort to force and violence'. This decision established that in order to be punishable, inflammatory speech must be likely to incite 'imminent lawless action'.[24] These free speech commitments explain the initial reluctance of major US social media companies to take down even IS content. The American case is a complete outlier when considered in a global context, with such a permissive approach

[21] J M Berger and Jonathon Morgan, *The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter* (Washington, DC: Brookings, 2016), p. 9; Berger, *The Alt-Right Twitter Census*, p. 13.
[22] YouTube, 'Our Ongoing Work to Tackle Hate', blog post, 5 June 2019, <https://youtube.googleblog.com/2019/06/ourongoing-work-to-tackle-hate.html>, accessed 20 January 2020.
[23] Elizabeth Dwoskin, 'YouTube's Arbitrary Standards: Stars Keep Making Money Even After Breaking the Rules', *Washington Post*, 9 August 2019. It is worth noting that YouTube has since expanded its anti-harassment policy – it is, however, too soon to tell what the impact of this latest update will be. For more, see Casey Newton, 'YouTube Expands Anti-Harassment Policy to Include All Creators and Public Figures', *The Verge*, 11 December 2019.
[24] Noto vs. United States, 367 US 290, 297-298 (1961), quoted in Brandenburg vs. Ohio, 395 US 444, 448 (1969).

to hateful speech not shared by other Western democracies. Facebook made the following comparison in an official 2017 blog post:

> In Germany, for example, laws forbid incitement to hatred; you could find yourself the subject of a police raid if you post such content online. In the US, on the other hand, even the most vile kinds of speech are legally protected under the US Constitution.[25]

First Amendment values have significantly impacted the response to hate speech and right-wing extremist content of major US social media companies to date. Even though, as private companies, major social media platforms have a relatively free hand as regards the restrictions contained in their terms of service (due to the provisions of Section 230 of the US Communications Decency Act),[26] they are nonetheless at all times mindful of opening themselves up to accusations of censorship and contravening the spirit, if not the letter, of the First Amendment. The companies' robust, years-long defence of First Amendment principles means that content moderation has therefore come to be viewed by some users as a free speech versus censorship issue and many, even outside of the US, can now be heard complaining about the contravention of their First Amendment rights. This is particularly true within the online scene of the extreme right, where this is a major rallying cry.[27] Again, there was no similar outcry regarding the free speech rights of IS activists and supporters.

Banished from most major platforms, IS does not have a lot of other high-profile and high-traffic platforms to rely on. Telegram – on which IS is still active – is an exception, despite a major purge by the platform, in conjunction with Europol, of IS-supporting accounts in November 2019.[28] In contrast, the increasingly inhospitable environment of major social media and other online platforms to extreme-right content resulted in US far-right activists establishing their own platforms that welcome, indeed encourage, just such activity. Established in 2016, Gab is currently the most prominent 'alt-tech'/'alt-right' platform.[29] Gab came to public attention when it emerged that Robert Bowers, the man who allegedly killed 11 people at a Pittsburgh synagogue in October 2018, had been an active user. Gab's founder is dismissive of the platform's role in radicalisation to terrorism.[30] In light of this, and in contrast to the example of IS, it seems unlikely that

---

[25] Richard Allen, 'Hard Questions: Who Should Decide What Is Hate Speech In An Online Global Community?',
Facebook Newsroom, 27 June 2017, <https://newsroom.fb.com/news/2017/06/hard-questions-hate-speech/>, accessed 20 January 2020.
[26] 47 US Code §230, 'Protection for Private Blocking and Screening of Offensive Material'. For more, see Annemarie Bridy, 'Leveraging CDA 230 to Counter Online Extremism', The George Washington University Program on Extremism, September 2019, <https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/Leveraging%20230%20to%20Counter%20 Online%20Extremism.pdf>, accessed 20 January 2020.
[27] Savvas Zannettou et al., 'What Is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber?', *WWW '18: Companion Proceedings of the The Web Conference*, pp. 1007–14. See also Avi Selk, 'How Deplatforming Became a Rallying Cry for Right-Wing Media Stars', *Washington Post*, 11 July 2019.
[28] Europol, 'Referral Action Day Against Islamic State Online Terrorist Propaganda', 22 November 2019, <https://www.europol.europa.eu/newsroom/news/referral-action-day-against-islamic-state-online-terrorist-propaganda>, accessed 20 January 2020.
[29] David Gilbert, 'Here's How Big Far Right Social Network Gab Has Actually Gotten', *VICE*, 16 August 2019.
[30] Craig Timberg et al., 'From Silicon Valley Elite to Social Media Hate: The Radicalization that Led to Gab', *Washington Post*, 31 October 2018; Gilbert, 'Here's How Big Far Right Social Network Gab Has Actually Gotten'.

small and fringe outfits will join the online fight against the extreme right.[31] On the contrary, as far as the founders and/or owners of fringe platforms are concerned, they are taking a principled and constitutionally endorsed stand in favour of free speech.

Conclusion

The extreme right cannot legitimately be compared to IS. It is a fast-changing 'scene' rather than a group and, while individuals involved in that scene have carried out terrorist attacks, very few groups have official terrorism designations as relatively few attackers could be confirmed as members or even supporters of formal groups. Instead, they can only be described as espousing common extreme right-wing talking points. While this does not entirely prevent the ability of online platforms to take down content (as it does not have to be terrorist in nature to be subject to removal), it is complicating in the present climate. Powerful actors – including heads of state, major political parties, some traditional media organisations and broad swathes of Western publics – identify with this content. This leaves social media companies open to charges of bias, which is almost exactly the opposite of what occurred in respect of their dealings with IS, when the same powerful actors pressured them to take decisive action to remove IS content from their platforms.

Even when online companies have acted against extreme-right content on their platforms, their cheerleaders are much better placed than IS to establish and maintain their own social media platforms. This is particularly true of extreme-right activists in the US, who can invoke strong First Amendment protections.

In conclusion, major social media companies have taken steps in recent years to stem the flow of hate content across their platforms. Greater pressure has, nonetheless, come to bear on them since the Christchurch attack and the spate of other right-wing extremist attacks that occurred in its wake. The Christchurch Call, for example, represents 'a commitment by Governments and tech companies to eliminate terrorist and violent extremist content online'.[32] In fact, concerted efforts were made by major – and a raft of minor – social media companies to remove all content directly related to the attack from their platforms.

Ultimately, the issue is not about the upload and circulation of extreme right-wing *terrorist* content on major social media platforms – this can be addressed in similar ways to how IS content was dealt with. Instead, the issue is more nuanced and contentious, and concerns the appropriateness of removing hostile extreme right-wing content from major platforms prior to attacks due to its potentially radicalising nature.

---

[31] This stands in contrast to the action taken by small platforms against IS content and activity (for example, the attempts by JustPasteIt to remove IS content from its platform). See Carmen Fishwick, 'How a Polish Student's Website Became an Isis Propaganda Tool', *The Guardian*, 15 August 2014.
[32] See Christchurch Call, <https://www.christchurchcall.com/>, accessed 20 January 2020.