DIGITAL JIHAD

Online Communication and Violent Extremism

edited by **Francesco Marone**
introduction by **Paolo Magri**

ISPI

DIGITAL JIHAD

Online Communication and Violent extremism

edited by Francesco Marone

ISPI

Via Alamanni, 11 – 20141 Milano – Italy
www.ledizioni.it
info@ledizioni.it

Digital Jihad. Online Communication and Violent Extremism
Edited by Francesco Marone
First edition: November 2019

This report is published with the support of the Italian Ministry of Foreign Affairs and International Cooperation, in accordance with Article 23- bis of the Decree of the President of the Italian Republic 18/1967. The opinions expressed are those of the authors. They do not reflect the opinions or views of the Italian Ministry of Foreign Affairs and International Cooperation.

Print ISBN 9788855261357
ePub ISBN 9788855261364
Pdf ISBN 9788855261371
DOI 10.14672/55261357

ISPI. Via Clerici, 5
20121, Milan
www.ispionline.it

Catalogue and reprints information: www.ledizioni.it

Table of Contents

7. Terrorist Content and the Social Media Ecosystem: The Role of Regulation

Patrick Bishop, Stuart Macdonald

Two months after the two attacks on mosques in Christchurch, New Zealand Prime Minister, Jacinda Ardern, and French President, Emmanuel Macron brought together Heads of State and Government and leaders from the tech sector to adopt the Christchurch Call. Among the commitments listed in the call was a pledge from the Governments to "Consider appropriate action to prevent the use of online services to disseminate terrorist and violent extremist content", including "Awareness-raising and capacity-building activities aimed at smaller online service providers" and "Regulatory or policy measures consistent with a free, open and secure internet and international human rights law"[1].

On the same day, a consortium of technology companies – including Amazon, Facebook, Google, Twitter and Microsoft – released a list of nine steps it would take to "address the abuse of technology to spread terrorist content"[2]. These included both individual actions (such as continued investment in technology to improve detection and removal of terrorist and violent extremist content) and collective actions (such as working across

[1] https://www.christchurchcall.com/call.html (last retrieved on 12 July 2019).

[2] GIFCT, Actions to Address the Abuse of Technology to Spread Terrorist and Violent Extremist Content, 15 May 2019 (last accessed 12 July 2019); Amazon, Microsoft, Google, Facebook, and twitter, Joint Statement in Support of Christchurch Call (last retrieved on 12 July 2019).

industry, governments, and NGOs to create crisis protocols).

The Christchurch Call is the latest in a number of efforts in recent years to require social media companies to do more to ensure that terrorist content is removed from their platforms. In 2018, the European Commission published a proposal for a new Regulation on preventing the dissemination of terrorist content online[3]. Aiming to balance the swift and effective detection and removal of terrorist content with the protection of human rights, Article 3 of the Regulation proposes the creation of a general duty for hosting services to "take appropriate, reasonable and proportionate actions […] against the dissemination of terrorist content and to protect users from terrorist content". Article 4 proposes the introduction of a removal order. This could be issued either administratively or judicially and would oblige the relevant platform to remove the content within one hour[4].

A year earlier, in 2017, Germany passed its Network Enforcement Act ("NetzDG"). This applies to all for-profit social media platforms with at least two million registered users in Germany. The NetzDG law requires platforms to remove or block obviously illegal content within 24 hours and to decide on all other complaints within one week[5]. Fines of up to €50 million can be imposed in cases involving systematic breaches of the law[6].

The introduction of a similar law has also been approved by the French National Assembly and is currently awaiting consideration by the Senate[7].

---

[3] European Commission, "Proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online", 2018/0331 (COD), 2018.

[4] Article 18 provides that member states should lay down the penalties for non-compliance.

[5] W. Echikson and O. Knodt, *Germany's NetzDG: A key test for combatting online hate*, CEPS Research Report, 2018.

[6] Section 4(2) of NetzDG invokes section 30(2) of the Act on Regulatory Offences, which states that the maximum regulatory fine may be multiplied by ten.

[7] Agence France-Presse, "France online hate speech law to force social media

In the UK, meanwhile, in April 2019 the Government proposed the creation of a new independent regulatory body in its *Online Harms* white paper. The regulator would enforce a new statutory duty of care on relevant companies to take reasonable steps to keep their users safe and tackle illegal and harmful activity on their services, with a range of enforcement options including the imposition of fines[8].

Against this backdrop, the starting point for this chapter is not *whether* regulatory measures should be imposed that require social media companies to do more to remove terrorist content from their platforms, but *what form* such measures should take. The chapter will argue, first, that a diverse regulatory toolkit is essential. There is no one-size-fits-all regulatory intervention. Whilst public discourse has tended to focus on the imposition of fines, other measures such as capacity-building, removal orders and the disruption of business activities (e.g. removal from search engine results and ISP blocking) are also necessary.

Second, the chapter will argue that efforts to regulate social media companies must be responsive to a range of factors, including the company's size and the extent of its engagement with the regulator. The benefits of responsive regulation are well-established and have been discussed extensively in academic literature[9].

Third, in order to ensure responsivity, the regulatory toolkit should be arranged in a pyramid structure, where each layer of the pyramid consists of sanctions of increasing severity. The underlying logic is that regulated entities will be more likely to engage with the less draconian interventions at the base of the pyramid when faced with the prospect of escalation and increasingly severe penalties.

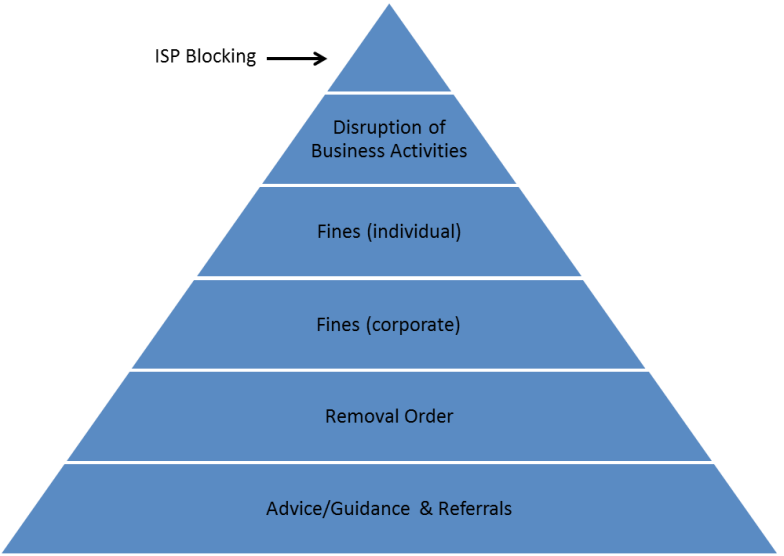
---

sites to act quickly", *The Guardian*, 9 July 2019.

[8] HM Government, *Online Harms White Paper*, The Stationery Office, 2019.

[9] I. Ayres and J. Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate*, Oxford, Oxford University Press, 1992; J. Braithwaite, *Restorative Justice and Responsive Regulation*, New York, Oxford University Press.

The chapter accordingly proposes the following enforcement pyramid:



At the bottom are advice, guidance and referrals. These strategies seek to support companies' efforts to self-regulate their platforms. For companies that fail to do so, the next layers of the pyramid are removal orders and, in the event of breach of a removal order, fines. At the top of the pyramid are disruption of business activities and, as a last resort, ISP blocking.

The chapter begins by explaining that efforts to remove online terrorist content must target the whole of the social media ecosystem, not just the social media giants. Having shown how terrorist groups exploit smaller platforms in order to ensure relatively stable access to their propaganda, the chapter then discusses three types of company in turn: the social media giants; smaller companies that lack the capacity to regulate their platforms effectively; and, smaller companies that lack the willingness to do so.

The Social Media Ecosystem

Before the Christchurch attacks, the attacker uploaded his manifesto to a range of smaller file-sharing sites (including MediaFire, ZippyShare and Solidfiles). Shortly before the first attack, he went onto Facebook, Twitter and 8chan and posted links to the copies of his manifesto available on these file-sharing sites. The post on 8chan also included a link to his Facebook profile, through which he livestreamed the attack[10]. Facebook has reported that the video was viewed fewer than 200 times during the live broadcast[11]. The first user report on the original video arrived 29 minutes after the video started, and 12 minutes after the live broadcast ended, by which time a user on 8chan had already posted a link to a copy of the video on a file-sharing site[12]. The video was subsequently shared on YouTube, as well as smaller platforms LiveLeak, BitChute and Kiwifarms, and as a downloadable file on Torrentz. Further links to the attack were re-shared on Facebook, Reddit, and 8chan. Whilst most of the smaller platforms reacted responsibly, some did not and still have active links to the video and manifesto[13].

Facebook has stated that, in the 24 hours after the attack, it blocked more than 1.2 million videos of the attack at upload[14]. A further 300,000 copies were removed after they were posted. One of the reasons why these additional copies were not detected by Facebook's image and video matching technology was the proliferation of different variants of the video: more than 800 "visually-distinct variants" were in circulation[15]. Some of these were the product of "a core community of bad actors working

[10] Tech Against Terrorism, "Analysis: New Zealand attack and the terrorist use of the internet", 26 March 2019 (last retrieved on 12 July 2019).
[11] G. Rosen, "A Further Update on New Zealand Terrorist Attack", *Facebook News*, 20 March 2019 (last retrived on 12 July 2019).
[12] Ibid.
[13] Tech Against Terrorism (2019).
[14] G. Rosen, A Further Update on New Zealand Terrorist Attack..., cit.
[15] Ibid.

together to continually re-upload edited versions of this video in ways designed to defeat our detection"[16].

The dissemination of the Christchurch attack video has similarities with IS (Islamic State)'s propaganda dissemination strategy. In the past five years IS's presence on Twitter has greatly diminished and a migration to Telegram has occurred[17]. Whilst Telegram's suite of features is used by IS supporters to interact and communicate, to distribute joinlinks to other groups and channels and to provide instructional materials, by far the most common function is the distribution of core IS media and, in particular, other pro-IS materials (regardless of their origin)[18]. As well as "using Telegram's file-sharing features to disseminate content internally, IS sympathizers on Telegram use external file-sharing sites to ensure IS content remains on the internet and resilient to takedowns"[19]. Dozens of unique URLs to a single piece of pro-IS material on different file-sharing sites are distributed using Telegram channels and groups. The URLs are then shared on Twitter, Facebook, and other mainstream social media platforms[20]. This separation of the content producers, disseminators, and consumers from the material itself bolsters IS dissemination networks against the effects of takedowns by ensuring that, even if content is removed from one site, stable access exists to others[21]. File-sharing platforms are thus utilised as "communication black-boxes" to "enable the rapid redistribution of content even under conditions of drastic policing

[16] Ibid.

[17] M. Conway et al., *Disrupting Daesh: Measuring Takedown of Online Terrorist Material and its Impacts*, VOX-Pol Network of Excellence, 2017.

[18] B. Clifford and H. Powell, *Encrypted Extremism: Inside the English-Speaking Islamic State Ecosystem on Telegram*, Washington, DC, George Washington University Program on Extremism, 2019.

[19] Ibid., p. 24.

[20] S. Macdonald, D. Grinnell, A. Kinzel, and N. Lorenzo-Dus, "Daesh, Twitter and the Social Media Ecosystem: A Study of Outlinks Contained in Tweets Mentioning Rumiyah", *The RUSI Journal*, 2019

[21] S. Weirman and A. Alexander, "Hyperlinked Sympathizers: URLs and the Islamic State", *Studies in Conflict & Terrorism*, 18 April 2018.

and filtering"[22]. The result is a "fragmentation" of IS propaganda that makes these materials "less trackable by authorities" and results in a "relatively closed and stable digital propaganda ecosystem"[23].

The Social Media Giants

As the previous section showed, efforts to remove online terrorist content must address the whole social media ecosystem. This section discusses the application of the regulatory pyramid outlined above to the social media giants. The next sections then turn to smaller platforms.

Self-regulatory measures

Given the sheer volume of content posted and uploaded to social media every day, the use of technology to identify and remove terrorist content is essential. Some progress has already been made in this respect. Facebook utilises image matching (so that, if someone tries to upload a photo or video that matches a photo or video that has previously been identified as terrorist, they are prevented from doing so), language understanding (analysing text that has been removed for praising or supporting terrorist organisations in order to develop text-based signals that can go into machine learning algorithms to detect similar future posts), removing terrorist clusters (using algorithms to work outwards from pages, groups, posts or profiles that have been identified as supporting terrorism, employing signals such as whether an account is friends with a high number of accounts that have been disabled for terrorism) and tackling recidivism

[22] T.E. Mitew and A. Shehabat, "Black-boxing the Black Flag: Anonymous Sharing Platforms and ISIS Content Distribution Tactics", *Perspectives on Terrorism*, vol. 12, no. 1, 2018, pp. 81-99, 84, 97.

[23] L. Bindner and R. Gluck, *Trends in Islamic State's Online Propaganda: Shorter Longevity, Wider Dissemination of Content*, ICCT Perspective, International Centre for Counter-Terrorism, The Hague, 2018.

(detecting new accounts created by repeat offenders)[24].

In the first quarter of 2019, Facebook removed 6.4 million items of terrorist propaganda. In 2018 it removed 19 million[25]. Twitter also uses its own technology to identify accounts promoting terrorism, as does YouTube[26]. From 1 August 2015 to the end of 2018, Twitter suspended a total of 1,582,026 accounts for the promotion of terrorism[27]. From September 2018 to March 2019, YouTube removed 149,980 videos for the promotion of violence and violent extremism[28]. Across all three platforms, referrals from users, law enforcement and governments are responsible for only a small minority of suspensions and take-downs; the vast majority of violations are detected by technology[29].

In addition, in order to try and prevent terrorists jumping from one platform to another, in 2017 Facebook, Google, Microsoft and Twitter founded the Global Internet Forum to Counter Terrorism (GIFCT). Members of the GIFCT collaborate to disrupt terrorist exploitation of their platforms. This includes a shared database of hashes (i.e. unique digital fingerprints). When a violent terrorist image or terrorist recruitment video is removed from a member company's platform, its hash is shared with other GIFCT members, enabling them to identify and remove it – or block it before it has even been posted.

---

[24] M. Bickert and B. Fishman, "Hard Questions: How We Counter Terrorism", *Facebook News*, 15 June 2017 (last retrieved on 14 July 2019).

[25] G. Rosen, "An Update on How We Are Doing At Enforcing Our Community Standards", *Facebook News*, 23 May 2019 (last retrieved 14 July 2019).

[26] V. Gadde, "Key data and insights from our 14th Twitter Transparency Report", *Twitter Public Policy Blog*, 9 May 2019 (last retrieved on 14 July 2019); S. Wojcicki, "Expanding our work against abuse of our platform", *YouTube Official Blog*, 4 December 2017 (last retrieved 14 July 2019).

[27] Figures taken from Twitter's biannual transparency reports.

[28] Figures taken from the "YouTube Community Guidelines" section of Google's transparency reports, available at https://transparencyreport.google.com/youtube-policy/removals?hl=en (last retrieved on 14 July 2019).

[29] M. Bickert, "Hard Questions: What Are We Doing to Stay Ahead of Terrorists?", *Facebook News*, 8 November 2018 (last retrieved on 14 July 2019); V. Gadde (2019).

There are currently 14 GIFCT members and over 200,000 hashes in the shared database[30].

Removal orders and fines

Whilst some progress has been made, there nonetheless remains a conviction that reliance on self-regulation is insufficient and that the biggest social media companies should be doing more[31]. According to the European Commission, for example, "the scale and pace of progress among hosting service providers as a whole is not sufficient to adequately address this problem"[32]. This has led to the various regulatory measures and proposals outlined in the introduction. At the same time, however, attempts to impose sanctions on the biggest social media companies raise a number of difficult issues.

The use of fines and other financial penalties (whether imposed administratively or judicially) is an almost ubiquitous feature of regulatory regimes. As Germany's NetzDG law illustrates, fines may be utilised for a number of reasons. These include: failure to respond to a removal order within the allotted timeframe; failure to comply with reporting/transparency mechanisms; and failure to ensure an effective complaints mechanism. In addition to the direct economic impact, the reputational damage associated with a fine may provide an additional incentive for compliance[33]. In particular, it may affect the platform's attractiveness as an advertising space, as highlighted by the UK Parliament's Intelligence and Security Committee:

---

[30] See https://gifct.org/joint-tech-innovation/ (last accessed 14 July 2019).

[31] See, for example, the evidence session before the UK Parliament's Home Affairs Committee, at which the Chair of the Committee stated to representatives from Facebook, YouTube and Twitter: "It seems to me that time and again you are simply not keeping up with the scale of the problem, the scale of criminal and terrorist activity, and doing the things that we all, as communities across the world, need you to be doing" (Home Affairs Committee, "Oral evidence: Hate crime and its violent consequences", HC 683 24 April 2019 Q904).

[32] European Commission (2018).

[33] ICF Consulting Services Ltd, *Research into Online Platforms' Operating Models and Management of Online Harms*, 2019 (last retrieved on 16 July 2019).

> When there was a social media backlash against companies whose adverts appeared alongside extremist videos on YouTube, those companies had little choice but temporarily to stop advertising on YouTube. More recently, Unilever announced that it is considering withdrawing its business from companies that are not doing more to provide "responsible digital infrastructure[34].

At the same time, the imposition of fines may not be straightforward. For a start, a company may not be willing to pay and, if it is registered outside of jurisdiction and does not have any physical assets within jurisdiction, it may not be possible to enforce it. In terms of the social media giants, however, this is less of a problem. Facebook, Google, Microsoft and Twitter all have physical headquarters within the EU and, as such, member states may enforce fines via a system of mutual recognition[35].

A more significant limitation stems from these companies' sheer economic strength. Any fine may simply be absorbed as an additional cost of doing business. One potential solution here is to base the level of fines on the financial strength of the platform. Under the General Data Protection Regulation, for example, the most serious infringements may incur a fine of up to €20 million or 4 per cent of the company's total worldwide turnover in the preceding financial year, whichever is higher[36].

In traditional business models, there is a danger that the cost of substantial fines will ultimately be borne by consumers. As one commentator has remarked, "when the corporation catches a cold, some else sneezes"[37]. However, since most social media

---

[34] Intelligence and Security Committee of Parliament, *The 2017 Attacks: What needs to change?*, 2018 (last retrieved on 19 August 2019).

[35] Council Framework Decision 2005/214/JHA of 24 February 2005 on the application of the principle of mutual recognition to financial penalties.

[36] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), art.83(5).

[37] J. Coffee, "No soul to damn: no body to kick: an unscandalized inquiry into the problem of corporate Punishment", *Michigan Law Review*, vol. 79, no. 386, 1980, p. 389.

companies provide services to users free of charge, a different dynamic applies. Perhaps the cost of fines will be passed on to advertisers, which could lead to reduced investment in advertising on the platform. Or perhaps the impact of a substantial fine will be mitigated by increasing the volume of advertising on the platform, which could spoil the user experience and impact the number of active users. Either way, the possibility of a substantial fine provides an incentive to the company to ensure compliance with the regulatory regime in the first place.

There is also the possibility of imposing personal liability on senior management. Individual liability in the context of corporate transgressions is a growing trend[38]. The UK's recent *Online Harms* white paper considers the possibility of senior management liability for major breaches of the proposed statutory duty of care but acknowledged a number of challenges, including identifying the roles to which liability might attach[39]. Within the UK, the Data Protection Act 2018 permits individual liability in cases where an offence has been committed by a corporation with the consent or connivance of a director, manager, secretary or similar officer, or is attributable to the neglect of one of these[40]. However, even if a person with a suitably senior role resides within jurisdiction, the requirement to demonstrate consent, neglect, etc. is a potentially difficult hurdle to overcome. This might be particularly challenging in the case of social media giants with complex management structures.

Disruption of business activities and ISP blocking

In extreme cases it may be necessary to resort to the most draconian enforcement options at the top of the regulatory pyramid: disruption of business activities and ISP blocking. Both of these possibilities are considered by the UK's *Online Harms*

[38] M. Nietsch, "Corporate illegal conduct and directors' liability: an approach to personal accountability for violations of corporate legal compliance", *Journal of Corporate Law Studies*, vol. 18, no. 1, 2018, pp. 151-184.

[39] Above, no. 8, p. 60.

[40] Data Protection Act 2018, s.198(1).

white paper. The former would include requiring third parties to withdraw services from the transgressing company, including removal from search results and app stores, and the cancelation of a range of ancillary services such as domain name registration and payment processing. Such measures might restrict the future growth of platforms, although in the case of the social media giants – who already have extremely large numbers of existing registered users – the impact may be minimal.

ISP blocking represents an option of last resort and raises some significant issues. Social media has become an integral aspect of people's everyday lives. Blocking access to the biggest platforms would cause public outcry and would have a significant socio-economic impact. It would also represent a prior restraint on speech, something which is generally seen as antithetical to liberal democratic traditions[41]. Whilst prior restraints may sometimes be consistent with the European Convention on Human Rights[42], it is questionable whether blocking an entire platform will be deemed proportionate when the vast majority of activity on the platform is lawful. There are also additional technological challenges, not least the ability of users to circumvent measures such as ISP blocking. This is a significant issue given that "we are entering an online environment in which the knowledge barrier for using technologies such as VPNs and TOR has never been lower"[43].

Smaller Companies That Lack Capacity

Governments may be less concerned about the impact of terrorist content published on small- or micro-platforms because, by definition, such platforms have relatively limited reach. The NetzDG law, for example, only applies to platforms with

[41] E. Barendt, *Freedom of Speech*, Oxford, Oxford University Press, 2007.
[42] *Observer and Guardian v UK*, 14 EHRR 153, 1992.
[43] CYTREC, *Response to the Online Harms White Paper* , 2019 (last retrieved on 16 July 2019).

over two million registered users in Germany. Yet, as explained above, the social media giants form just one part of a wider ecosystem. It is through the exploitation of smaller platforms that terrorist groups have managed to bolster resilience against take-downs and ensure stable access to their propaganda. Moreover, the more effective the regulation of the biggest platforms, the more likely terrorist migration to smaller platforms becomes.

In terms of these smaller platforms, the challenge is often not a lack of willingness but a lack of capacity. A well-known example is Justpaste.it. Justpaste.it is a free content-sharing service that allows content to be posted within seconds with no registration required. Owned by Mariusz Zurawek, who runs the site out of his home in Poland, the content posted on Justpaste.it began to include IS propaganda in early 2014. By March 2015, Zurawek estimated that he had removed up to 2,000 posts at the request of London Metropolitan Police[44]. Since then he has received a large volume of take-down requests from all over the world. This poses challenges in terms of identifying what content is legal and responding to take-down requests in other languages, as well as capacity and resources[45]. So, whilst any potential regulatory framework ought to apply to smaller platforms, it must also be proportionate and mindful of such platforms' relatively limited resources.

When dealing with such platforms, the central role of any regulator is to enable self-regulation by the provision of advice and guidance. In the UK's *Online Harms* white paper, for example, the functions of the proposed regulator include: the publication of codes of practice; the establishment of a transparency, trust and accountability framework; and, the provision

[44] S. Stalinsky and R. Sosnow, *The Jihadi Cycle On Content-Sharing Web Services 2009-2016 And The Case Of Justpaste.it: Favored By ISIS, Al-Qaeda, And Other Jihadis For Posting Content And Sharing It On Twitter - Jihadis Move To Their Own Platforms (Manbar, Nashir, Alors.Ninja) But Then Return To Justpaste.it*, MEMRI Inquiry & Analysis Series No. 1255, 6 June 2016 (last retrieved on 15 July 2019).

[45] Tech Against Terrorism, "UK Launch of Tech Against Terrorism at Chatham House", 12 July 2017 (last retrieved on 15 July 2019).

of support to start-ups and SMEs to help them fulfil their legal obligations in a proportionate and effective manner[46]. One of the objectives of the GIFCT is also to build the capacity of smaller platforms by knowledge-sharing. To this end, it has collaborated with the U.N.-mandated Tech Against Terrorism initiative to launch a knowledge-sharing platform[47]. This offers various resources including a list of terrorist groups and individuals on the U.N. sanctions list, recommendations for model terms of service and model guidelines for transparency reports.

Smaller platforms are also likely to have limited technological capacity. Here, the *Online Harms* white paper proposes equipping a regulator with powers to facilitate the sharing of technological solutions. It offers the example of a hackathon, attended by leading tech firms and hosted by the UK's Home Secretary and Microsoft, which commenced the development of a tool to identify online grooming. When complete the tool will be licensed free of charge to smaller and medium companies worldwide[48].

As mentioned above, GIFCT members have already collaborated to develop a shared hash database. Yet more could be done. GIFCT members are reportedly experimenting with URL sharing[49]. This is welcome, given the important role URLs play in propaganda dissemination strategies. Since botnets have also been found to play a significant role in efforts to disseminate terrorist propaganda, GIFCT members should also develop shared automated systems that use behavioural (as opposed to content-based) cues to block terrorist content (such as abnormal posting volume or the use of trending hashtags to

[46] Above, no. 8, para. 5.2.

[47] GIFCT, "Global Internet Forum to Counter Terrorism: an update on our efforts to use technology, support smaller companies and fund research to fight terrorism online", 18 June 2018 (last retrieved on 15 July 2019).

[48] UK Government, "New tool developed to tackle online child grooming", 13 November 2018 (last retrieved on 17 July).

[49] G. Rosen, A Further Update on New Zealand Terrorist Attack..., cit.

gain attention)[50]. This is important, given that many smaller platforms rely exclusively on humans to identify and remove terrorist content[51].

There is also a pressing need to expand membership of the GIFCT, to ensure access to such initiatives. One study found evidence of more than 330 different platforms being used by terrorist groups since 2016, with 25 of the top 50 most-used platforms being small- or micro-platforms[52]. In comparison, there are just fourteen GIFCT members. Many smaller platforms lack the capacity required to fulfil the GIFCT membership criteria. These criteria include: terms of service that include content standards; regular, public data transparency; a public commitment to human rights; and, support for civil society organisations challenging violent extremism[53]. Here, the provision of advice and guidance by a regulator, along the lines envisaged by the UK's *Online Harms* white paper, could have significant value.

Lastly, it should be noted that, whilst the role of a regulator in the case of smaller companies that are willing to cooperate would naturally be focused on interventions located at the base of the regulatory pyramid, a responsive approach would nonetheless require escalation to stricter interventions if self-regulatory strategies do not achieve the desired outcomes. Where this occurs, a truly responsive approach would require an element of proportionality, for example, by allowing a longer period to comply with a removal order than would be case with a social media giant.

---

[50] S. Macdonald, D. Grinnell, A. Kinzel, and N. Lorenzo-Dus, *A Study of Outlinks Contained in Tweets Mentioning Rumiyah*, Global Research Network on Terrorism and Technology, Paper no. 2, 2019 (last retrieved on 15 July 2019).

[51] I. van der Vegt, P. Gill, P., S. Macdonald, and B. Kleinberg, *Shedding Light on Terrorist and Extremist Content Removal*, Global Research Network on Terrorism and Technology, Paper no. 3, 2019 (last retrieved on 15 July 2019).

[52] Tech Against Terrorism, "ISIS use of smaller platforms and the DWeb to share terrorist content", 29 April 2019 (last retrieved on 15 July 2019).

[53] See https://gifct.org/members/ (last retrieved on 15 July 2019).

Smaller, Uncooperative Companies

As the video of the Christchurch attack illustrates, there may be some smaller platforms that are unwilling to remove terrorist content. From a regulatory perspective, these companies pose different challenges. There is little reason to provide advice and guidance on regulatory compliance if the platform in question does not accept the desirability of removing terrorist content in the first place. Facilitation strategies require a predisposition towards compliance and the broad objectives of the regulatory framework.

The use of fines in cases involving such companies may also be problematic. The enforcement of any financial penalties will depend on the nature of the jurisdiction where the platform is registered and senior management is domiciled. It is not difficult to imagine a scenario in which a platform is based in an uncooperative country and, in the absence of a reciprocal agreement with that country, any fines imposed are unenforceable.

Given the limited efficacy of traditional regulatory sanctions in cases involving uncooperative platforms, another enforcement strategy could be to force companies that provide supporting services to take responsibility. ISPs could, for example, be pressured to raise prices on such companies, provide them with a lower quality of service, or deny them service altogether. Following the October 2018 Pittsburgh Synagogue shooting, GoDaddy refused to be further associated with the controversial social media platform Gab, forcing it to find another domain provider[54]. Similarly, following the August 2019 shooting in El Paso, Texas, Cloudflare announced that it would no longer offer the message board 8chan protection from distributed denial of service attacks[55]. Such action might cause the

[54] Far-right social network Gab goes offline after GoDaddy tells it to find another domain registrar https://techcrunch.com/2018/10/28/far-right-social-network-gab-goes-offline-after-godaddy-tells-it-to-find-another-domain-registrar/ (last retreived on 19 July 2019).

[55] J. Taylor and J.C. Wong, "Cloudflare cuts off far-right message board 8chan

impugned company to change their behaviour or force them to seek another ISP, which would be a significant cost for a smaller company. The ultimate threat of ISP blocking should also remain in the regulator's armoury as a way of incentivising engagement with less severe interventions.

Conclusion

This chapter has identified three requirements for efforts to require social media companies to do more to remove terrorist content from their platforms to be effective. First, a diverse regulatory toolkit is required. The chapter has discussed a range of regulatory interventions, including advice and guidance, removal orders and fines, and disruption of business activity and ISP blocking. Each of these has value in certain contexts and limitations in others. Second, the chapter has shown the importance of responsivity to a range of factors, including: the nature of the relevant platform and its position within the social media ecosystem; the degree of company engagement with the regulator; and, the extent to which the conduct of companies enhances or undermines the overall rationale of the framework. Third, to achieve this responsivity, the various interventions in the toolkit should be organised in a regulatory pyramid.

Even the most carefully designed regulatory framework will not lead to perfect compliance. Some problems are likely to persist, particularly in respect of a regulatory framework that is limited to a single jurisdiction. Terrorist groups may migrate to platforms or jurisdictions which are relatively unregulated. Some regulatory tools, such as ISP blocking, may also be circumvented using such means as VPNs and the TOR browser, which enable access to platforms that are blocked in the user's home country but accessible in others. The obvious solution to these challenges is to adopt a global approach – but obtaining

after El Paso shooting", *The Guardian*, 5 August 2019 (last aretrieved on 23 August 2019).

international agreement is beset with difficulties. The prospects for a regional approach may be more favourable; the Council of Europe Convention on Cybercrime has not only been ratified by countries beyond Europe but has also influenced the design of cybercrime legislation in a number of non-signatory states[56]. Ultimately, however, the fact that a regulatory regime will not achieve perfect compliance is not a reason to not enact the regime in the first place. As Berger and Morgan state in their study of the disruption of IS activity on Twitter, "The consequences of neglecting to weed a garden are obvious, even though weeds will always return"[57].

[56] J. Clough, "A world of difference: the Budapest Convention on Cybercrime and the challenges of harmonisation", *Monash University Law Review*, vol. 40, no. 3, 2014, pp. 698-736.

[57] J.M. Berger and J. Morgan, *The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter*, Analysis Paper no. 20, Brookings Institution, Washington, DC, 2015, p. 56.

The Authors

Patrick Bishop is a Senior Lecturer in Law at the Hilary Rodham Clinton School of Law, Swansea University, and a member of the university's Cyber Threats Research Centre. Patrick's research interests include the regulation of cyberspace and the use of science in legal decision making. He a contributor to the Routledge *Handbook of Technology, Crime and Justice* (2017) and the Palgrave *Handbook of International Cybercrime and Cyberdeviance* (2019). Patrick's teaching includes Cybercrime at undergraduate level and Crime in Cyberspace on Swansea's MA in Cybercrime and Terrorism.

Matteo Colombo is an ISPI Associate Research Fellow in the MENA Centre, and a PhD Candidate in Political Studies at the University of Milan (NASP Consortium), with a thesis on the Twitter discourse about IS in Arabic. He holds a bachelor degree in Philosophy from the Catholic University of Milan, a Master in International Relations from Oxford Brookes and a Master in Middle East Politics from SOAS. He also holds a Master Degree in journalism at IULM University in Milan. He is a researcher, professional journalist and infographic designer. Before joining ISPI, he worked as a correspondent from Cairo for TgCom24, where he also studied modern standard Arabic. His main interests are in social dynamics, social media, jihadism, Arab politics, Sectarianism and the Italian Foreign Policy.

Ali Fisher is Explorer of Extreme Realms at Human Cognition, where he is part of an interdisciplinary team that works with governments and tech companies to track, analyze and disrupt the use of the internet by terrorist groups. Ali previously directed Mappa Mundi Consulting and the cultural relations thinktank, Counterpoint. He worked as associate director of Digital Media Research at Intermedia and as lecturer in International Relations at Exeter University. Ali received his Ph.D. at the University of Birmingham in 2006. His books include *Collaborative Public Diplomacy* (2012), *The Connective Mindshift* (2013), and *Trails of Engagement* (2010), Fisher's CPD Research Fellowship project at University of Southern California is titled *Netwar in Cyberia: Decoding the Media Mujahedeen and the Jihadist Swarmcast*.

Marco Lombardi, Director of ITSTIME (Italian Team for Security, Terroristic Issues & managing Emergencies – www.itstime.it) research center, is full professor at the Catholic University of Sacred Heart, where he teaches Crisis management and risk communication, Mass Communication, Sociology, Intelligence and counter terrorism, Security policies. He is the director the Department of Sociology and of the School of Journalism, member of the scientific board of the School of Doctorate and of the master in Cultural Diplomacy. He managed several EU founded research projects mainly focused on terrorism, security and crisis management. He co-operates with different institutional agencies on security both at national and international level. He is member of the Governmental Commission on Counter Radicalization and of the Strategic Policy Committee of the Italian Ministry of Foreign Affairs and International Cooperation.

Stuart Macdonald is Professor of Law at the Hillary Rodham Clinton School of Law, Swansea University. He is Director of the University's Cyber Threats Research Centre and Co-Director of the CHERISH Digital Economy Centre. Stuart's

research focuses on terrorist's use of the internet. He is the lead organiser of the biennial Terrorism and Social Media (TASM) conference, a member of Europol's Advisory Network on terrorism and propaganda and co-ordinates Swansea University's contribution to the Global Research Network on Terrorism and Technology. In 2016/17 he was also the holder of a Fulbright Cyber Security Award.

Francesco Marone, Ph.D., is a Research Fellow at the ISPI Center on Radicalization and International Terrorism. A social scientist by training, he is currently an Adjunct Lecturer in International Politics at the University of Pavia, Italy. Moreover, Francesco is an Associate Fellow of the International Centre for Counter-Terrorism – The Hague (ICCT), Netherlands, and a Fellow of the Program on Extremism at George Washington University, USA. His research interests include terrorism and counter-terrorism, violent extremism, and national security. He is the author of several publications in the field of radicalisation and international terrorism.

Valerio Mazzoni is a terrorism analyst for the Italian consultancy IFI Advisory, which has recently opened a monitoring unit on extremist online activities. He collaborates with The European Eye on Radicalization Studies Centre for which he has also led a research on far-right extremist online activities. He is also an author for the Italian geopolitical newspaper *Babilon Magazine* and for the website *Il Caffè Geopolitico*.

Daniele Plebani is Senior Researcher at the Italian Team for Security, Terroristic Issues & Managing Emergencies (ITSTIME) at the Università Cattolica del Sacro Cuore. His research focuses on the evolving equilibriums of the wider Middle East, cross-ideological tactics and propaganda of major terrorist organizations as well as on cultural heritage as an asset for stabilization in critical areas. He collaborated with CeSI-Università Cattolica on a project concerning cultural heritage

in Ethiopia and within the UNESCO framework for an initiative aimed at valorising the archaeological assets of Yazd in the Islamic Republic of Iran. He is member of the European Expert Network on Terrorism issues (EENeT) and his analyses were published by GNOSIS, Security, Terrorism & Society and ITSTIME.

Nico Prucha is Chief Content Curator at Human Cognition, responsible for curating, collecting and assessing mainly Arabic language videos, audios and writings published by Sunni extremist groups. He is a lecturer at the University of Vienna where he finished his doctorate on Sunni jihadist online operations analyzing their Arabic language output, in 2015. He is the author of Sawt al-Jihad al-Qaida's first electronic magazine (in German, published in 2010) and is currently working on two book manuscripts.

Manuel R. Torres Soriano is Professor of Political Science at Pablo de Olavide University of Sevilla (Spain) and Director of the Diploma in Analysis on Jihadist Terrorism, Insurgencies and Radical Movements of this university. He was a Visiting Fellow at Stanford University, Johns Hopkins University, King's College of London, London School of Economics and Harvard University. He is the author of the books *The Echo of Terror: Ideology and Propaganda in Jihadist Terrorism* (2009), *Al Andalus 2.0. The Cyber-jihad against Spain* (2014) and *#Disinformation. Power and Manipulation in Digital Era* (2019) (in Spanish). He is a member of the European Counter Terrorism Centre (ECTC) Advisory Network on terrorism and propaganda (EUROPOL).