Global Network
on Extremism & Technology

Conspiracy Theories, Radicalisation and Digital Media

Daniel Allington

The author of this report is
Daniel Allington

The Global Network on Extremism and
Technology (GNET) is an academic research
initiative backed by the Global Internet Forum to
Counter Terrorism (GIFCT), an independent but
industry-funded initiative for better understanding,
and counteracting, terrorist use of technology.
GNET is convened and led by the International
Centre for the Study of Radicalisation (ICSR),
an academic research centre based within the
Department of War Studies at King's College
London. The views and conclusions contained in
this document are those of the authors and should
not be interpreted as representing those, either
expressed or implied, of GIFCT, GNET or ICSR.

Executive summary

The purpose of this report is to explore the role that conspiracy theories, especially as disseminated through social media, may play in the process of radicalisation, and to make recommendations about how to minimise their occurrence.

As it will show, there is clear evidence:

- That conspiracy theories are disseminated through social networking and media sharing platforms

- That conspiracy theories have historically played an important role in radicalisation, terrorism, persecution and genocide

- That belief in conspiracy theories is psychologically associated with bigotry, extremism and willingness to break the law

- That the perpetrators and alleged perpetrators of many recent mass shooting events were motivated by belief in conspiracy theories

- That conspiracy theories have played a key role in recent political violence in the USA, including the insurrection of 6 January 2021

- That actions taken by social networking and media sharing platforms are inadequate to solve the problems associated with conspiracy theories, in part because the platforms themselves are designed in a way that serves to nurture and protect conspiracy beliefs

The report will conclude by suggesting that a cultural change is required in terms of how social networking and media sharing platforms understand their role. The tendency has been for them to view themselves as neutral spaces through which speakers are able to reach an audience (except under exceptional circumstances leading to the removal of this privilege), and to justify this self-conception through a misreading of the principle of freedom of speech. However, in the internet of today, value is increasingly attached not to platforms that facilitate an undifferentiated free-for-all but to platforms that provide *high quality* content, whether on a commercial basis (e.g. Disney+) or on a non-commercial basis (e.g. Wikipedia). Partnering with reputable content providers in order to promote high quality content at the expense of misinformation and conspiracy theories would in no way violate the principle of freedom of speech, and would indeed be likely to lead typical platform users to attach higher value to the platforms in question.

Contents

1 Introduction: What are 'Conspiracy Theories', and Why Should They Be Regarded as a Problem?

The term 'conspiracy theory' was coined by Karl Popper, who defined 'the conspiracy theory of society' as the false belief 'that institutions can be understood completely as the result of conscious design'.[1] Today, we tend to describe specific instances of this explanatory style as 'conspiracy theories'. Conspiracy theories are united by the claim 'not [only] that conspiracies happen, but that they are the motive force in history',[2] and require 'that there is an omnipotent secret group of people plotting to increase their own power at the expense of ordinary people'.[3] They constitute 'an explanation of politics [which] … purports to locate and identify the true loci of power … [among] conspirators, often referred to as a shadow or hidden government, [who] operate a concealed political system behind the visible one, whose functionaries are either ciphers or puppets'.[4] Such theories 'add up to an idea of the world in which the authorities, including those we elect, are systematically corrupt and untruthful'.[5] The associated mindset has been described as 'politically corrosive', potentially leading to scapegoating and violence as part of a withdrawal from democratic politics.[6]

The roots of conspiracist thinking are to be found in medieval superstitions that became secularised in the aftermath of the French Revolution. This point was first made by Joshua Trachtenberg while the Holocaust was at its height,[7] and was further developed by Norman Cohn,[8] who had encountered SS officers and their reading materials in the immediate aftermath of the Second World War. In pre-modern Europe, Jews were widely viewed as 'a league of

1   Popper, Karl, *Conjectures and refutations: the growth of scientific knowledge* (London: Routledge & Kegan Paul, 1969), p.168.
2   Byford, Jovan, *Conspiracy theories: a critical introduction* (Basingstoke: Palgrave Macmillan, 2011), p.34.
3   Yablokov, Ilya, *Fortress Russia: conspiracy theories in post-Soviet Russia* (Cambridge: Polity Press, 2018), p.1.
4   Barkun, Michael, *A culture of conspiracy: apocalyptic visions in contemporary America* (Los Angeles: University of California Press, 2003), p.178.
5   Aaronovitch, David, *Voodoo histories: how conspiracy theory has shaped modern history* (London: Vintage Books, 2010), p.5.
6   Muirhead, Russell, and Nancy L. Rosenblum, "Speaking truth to conspiracy: partisanship and trust." *Critical Review* 28:1 (2016), pp.63–88.
7   Trachtenberg, Joshua, *The devil and the Jews: the medieval conception of the Jew and its relation to modern antisemitism* (New Haven: Yale University Press, 1943).
8   Cohn, Norman, *Warrant for genocide: the myth of the Jewish world-conspiracy and the Protocols of the Elders of Zion* (London: Eyre & Spottiswoode, 1967).

sorcerers employed by Satan for the spiritual and physical ruination of Christendom',[9] and since the turn of the 18th and 19th centuries, they have been reimagined as 'a conspiratorial body set on ruining and then dominating the rest of mankind',[10] with the allegation of sorcery being replaced by the superficially more rational allegation of 'technological and economic mind control' through 'banks, mass media, government, [and] education'.[11] The 'blood libel' – the accusation that Jews conspire to murder children and drink their blood – is a closely related pre-modern myth that circulates in rationalised form even today.[12]

It is these forms of discourse that Herf sees as having been 'most important in fostering [the] radical, genocidal implications' of antisemitism under the Nazis.[13] However, Jewish people have not been the only victims of this dark tradition. The medieval European imagination conceived of heretics and witches in a very similar way to Jews,[14] including through the allegation of child-murder and child-eating,[15] and accusations of heresy and witchcraft were used for centuries as a tool of repression,[16] with barbaric punishments carried out as a public spectacle.[17] Moreover, the first targets of early conspiracy theorists Augustin Barruel and John Robison were not the Jews but the Freemasons and the (in reality, no longer extant) Illuminati,[18] and both the Nazis[19] and the Francoists[20] persecuted Freemasons harshly (although it should be noted that German Freemasons were able to escape persecution by leaving the organisation and aligning themselves with the Nazi regime).[21]

Given conspiracy theories' roots in pre-modern superstition, it seems paradoxical that they should be so closely associated with the internet. However, there exists a substantial body of research to indicate that social networking and media sharing platforms such as YouTube, Facebook, Twitter and Instagram serve as vectors for the dissemination of conspiracy beliefs and related forms of

---

9   *ibid.*

10 *ibid.*

11 Zukier, Henri, "The conspiratorial imperative: medieval Jewry in Western Europe." In *Changing conceptions of conspiracy*, edited by Carl F. Graumann and Serge Moscovici (New York / Berlin / Heidelberg / London / Paris / Tokyo: Springer-Verlag, 1987), p.95.

12 Hirsh, David, *Contemporary left antisemitism* (Abingdon: Routledge, 2017), p.206.

13 Herf, Jeffrey, *The Jewish enemy: Nazi propaganda during World War II and the Holocaust* (Cambridge, Massachusetts / London: The Belknap Press of Harvard University Press, 2006), p.10.

14 Trachtenberg, *The devil and the Jews*, p.215.

15 Cohn, Norman S., *Europe's inner demons: an enquiry inspired by the great witch-hunt,*(London: Chatto and Heinemann, 1975).

16 Caldwell Ames, Christine, *Righteous persecution: Inquisition, Dominicans, and Christianity in the Middle Ages* (Philadelphia: University of Pennsylvania Press, 2009); Murphy, Cullen, *God's jury: the Inquisition and the making of the modern world* (Boston / New York: Houghton Mifflin Harcourt, 2012).

17 Loewenstein, David, "Writing and the persecution of heretics in Henry VIII's England: the examinations of Anne Askew." In *Heresy, Literature, and Politics in Early Modern English Culture*, edited by David Loewenstein and John Marshall (Cambridge / New York / Melbourne / Madrid / Cape Town / Singapore / São Paulo / Delhi / Dubai / Tokyo: Cambridge University Press, 2006), pp.11–39.

18 Byford, *Conspiracy theories*, p.40.

19 Doney, Keith, "Freemasonry in France during the Nazi occupation and its rehabilitation after the end of the Second World War." PhD (University of Aston, 1993).

20 Ruiz, Julius, "Fighting the international conspiracy: the Francoist persecution of Freemasonry, 1936–1945." *Politics, Religion, and Ideology* 12:2 (2011), pp.179–96.

21 Thomas, Christopher Campbell, "Compass, square, and swastika: Freemasonry in the Third Reich." PhD (Texas A&M University, 2011).

misinformation.[22] Moreover, there are other popular online platforms where conspiracy theories have been found to circulate extensively, such as the comments sections of major newspapers.[23] Lastly, while conspiracy theories are partly a grassroots phenomenon, they are also the stock-in-trade of such online influencers as Alex Jones and David Icke. Perhaps more akin to scammers than to propagandists, these professional conspiracy theorists are able to extract large sums of money from their audiences through merchandising and online retail,[24] as well as through fundraising drives,[25] and have generated substantial advertising revenue for social networking and media sharing platforms.[26]

In context of the UK's counter-terrorism strategy, radicalisation is officially defined as 'the process by which a person comes to support terrorism and forms of extremism leading to terrorism' (where extremism is defined as 'vocal or active opposition' to values such as 'democracy, the rule of law, individual liberty, and mutual respect and tolerance of different faiths and beliefs').[27] Given the above observations, there is a clear risk that conspiracy theories may play a role in radicalisation so defined. Indeed, in 1970s Britain, one social psychologist found that conspiracy theories formed part of a sophisticated far-right radicalisation strategy in which potential recruits were invited to order literature that would fill in the explicitly

22 Pandey, A., N. Patni, M. Sing, A. Sood, and G. Singh, "YouTube as a source of information on the H1N1 influenza pandemic." *American Journal of Preventative Medicine* 38:3 (2010), pp.1–3; Buchanan, Rachel, and Robert D. Beckett, "Assessment of vaccination-related information for consumers available on Facebook." *Health Information and Libraries Journal* 31:3 (2014), pp.227–34; Oyeyemi, Sunday Oluwafemi, Elia Gabarron, and Rolf Wynn, "Ebola, Twitter, and misinformation: a dangerous combination?" *British Medical Journal* 349 (2014), p.6178; Seymour, Brittany, Rebekah Getman, Avinash Saraf, Lily H. Zhang, and Elsbeth Kalenderian, "When advocacy obscures accuracy online: digital pandemics of public health misinformation through an antiflouride case study." *American Journal of Public Health* 105 (2014), pp.517–23; Pathak, Ranjan, Dilli R. Poudel, Paras Karmacharya, Amrit Pathak, Madan Raj Aryal, Maryam Mahmood, and Andy Donata, "YouTube as a source of information on Ebola virus disease." *North American Journal of Medical Sciences* 7:7 (2015), pp.306–9; Chaslot, Guillaume, "How YouTube's A.I. boosts alternative facts: YouTube's recommendation A.I. is designed to maximize the time users spend online. Fiction often outperforms reality." *Medium* (31 March 2017), https://medium.com/@guillaumechaslot/how-youtubes-a-i-boosts-alternative-facts-3cc276f47cf7; Ortiz-Martínez, Yeimar, and Luisa F. Jiménez-Arcia, "Yellow fever outbreaks and Twitter: rumours and misinformation." *American Journal of Infection Control* 45 (2017), pp.815–16; Sharma, Megha, Kapil Yadav, Nikita Yadav, and Keith C. Ferdinand, "Zika virus pandemic – analysis of Facebook as a social media health information platform." *American Journal of Infection Control* 45:3 (2017), pp.301–2; Starbird, Kate, "Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter." Eleventh International AAAI Conference on Web and Social Media (2017); Bora, Kaustubh, Dulmoni Das, Bhupen Barman, and Probodh Borah, "Are Internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015–16 Zika virus pandemic." *Pathogens and Global Health* 112:6 (2018), pp.320–8; Broniatowski, David A., Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze, "Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate." *American Journal of Public Health* 108:10 (2018), pp.1378–84; Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira Jr., "Auditing radicalization pathways on YouTube." Woodstock '18: ACM symposium on neural gaze detection, Woodstock, NY, 3–5 June 2018; Basch, Corey H., Nicole Milano, and Grace C. Hillyer, "An assessment of fluoride related posts on Instagram." *Health Promotion Perspectives* 9:1 (2019), pp.85–8; Bovet, Alexandre, and Hernán Makse, "Influence of fake news in Twitter during the 2016 US presidential election." *Nature Communications* 10:7 (2019), pp.1–14; Ichau, Elke, Thomas Frissen, and Leen d'Haenens, "From #selfie to #edgy. hashtag networks and images associated with the hashtag #jews on Instagram." *Telematics and Informatics* 44 (2019); Massey, Philip M., Matthew D. Kearney, Michael K. Hauer, Preethi Selvan, Emmanuel Koku, and Amy E. Leader, "Dimensions of misinformation about the HPV vaccine on Instagram: content and network analysis of social media characteristics." *Journal of Medical Internet Research* 22:12 (2019); Allington, Daniel, and Tanvi Joshi, "'What others dare not say': an antisemitic conspiracy fantasy and Its YouTube audience." *Journal of Contemporary Antisemitism* 3:1 (2019): pp.35–53; Gruzd, Anatoliy, and Philip Mai, "Going viral: how a single tweet spawned a COVID-19 conspiracy theory on Twitter." *Big Data & Society* (2020); Kouzy, Ramez, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, and Khalil Baddour, "Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter." *Cureus* 12:3 (2020); Oi-Yee Li, Heidi, Adrian Bailey, David Huynh, and James Chan, "YouTube as a source of information on COVID-19: a pandemic of misinformation?" *BMJ Global Health* 5 (2020); Allington, Daniel, Beatriz Buarque, and Daniel Flores, "Antisemitic conspiracy fantasy in the age of digital media: three 'conspiracy theorists' and their YouTube audiences." *Language & Literature* (2021) published online first.
23 Wood, Michael J., and Karen M. Douglas, "Online communication as a window to conspiracist worldviews." *Frontiers in Psychology* 6 (2015).
24 Byford, *Conspiracy theories*, p.11.
25 Allington, Buarque, and Flores, "Antisemitic conspiracy fantasy."
26 CCDH, *#DeplatformIcke: How Big Tech powers and profits from David Icke's lies and hate, and why it must stop.* Center for Countering Digital Hate (London, 2020). https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_db8ff469f6914534ac02309bb488f948.pdf.
27 HM Government, *Prevent strategy. Cm 8092* (2011), pp.107–108.

racist details that publicly disseminated conspiracy theories left out,[28] and the subsequently published autobiography of a former leading neo-Nazi confirms that his own radicalisation almost exactly followed this pattern.[29] However, the radicalising potential of conspiracy theories is not necessarily limited to cases where they are disseminated by extremist organisations. The FBI, for example, has reported as follows:

> *The FBI assesses [that] anti-government, identity-based, and fringe political conspiracy theories very likely will emerge, spread, and evolve in the modern information marketplace over the near term, fostering anti-government sentiment, promoting racial and religious prejudice, increasing political tensions, and occasionally driving both groups and individuals to commit criminal or violent acts.*[30]

Direct calls to specific action are not typically made by leading conspiracy theorists, but their pronouncements often appear calculated to inspire feelings of grievance. For example, a book written by a conspiracy theorist frequently retweeted by US President Donald Trump begins by announcing its author's intention to arouse 'outrage at being lied to for so many years by the monstrous and well-oiled machine known as the Deep State'.[31] The question arises of whether the relatively unconstrained online circulation of such discourse in the absence of an explicit radical programme may nonetheless produce a general climate of undirected radicalism in which a proportion of individuals may spontaneously resort to acts of terrorism or perhaps even accept leadership from violent extremists. As the remainder of this report will show, the answer appears to be: Yes. But as the final section will argue, there is fortunately no reason why mainstream social networking and media sharing platforms should continue to perpetuate that state of affairs.

28   Billig, Michael, *Fascists: a social psychological view of the National Front* (London / New York / San Francisco: Academic Press, 1978), pp.172–3.
29   Ray, Hill, and Andrew Bell, *The other face of terror: inside Europe's neo-Nazi network* (London: Grafton Books, 1984), pp.28–36.
30   FBI Phoenix Field Office, *(U//LES) Anti-government, identity based, and fringe political conspiracy theories very likely to motivate some domestic extremists to commit criminal, sometimes violent activity.* (Federal Bureau of Investigation, 30 May 2019), p.5.
31   Stone, Jeremy, *History of the Deep State,* vol. 1 (Self-published, place of publication unknown, 2018), ch. 1.

2 Who Believes in Conspiracy Theories?

A wealth of psychological research has been carried out to discover the traits that predispose people to conspiracy beliefs. The classic study found that conspiracy beliefs are predicted by anomie (i.e. dissatisfaction with and rejection of society and its norms).[32] In one later study, conspiracy beliefs were found to be associated with anomie, authoritarian inclinations, low self-esteem and feelings of powerlessness,[33] while in another, endorsement of a new conspiracy theory invented by the researchers was found to be predicted by endorsement of existing conspiracy theories, belief in the paranormal and lower general intelligence.[34] The explanation for the repeated finding of correlation between unrelated (and, in some cases, contradictory) conspiracy beliefs is unclear.[35] In the classic study, Ted Goertzel influentially proposed that such beliefs correlate because they form a monological system,[36] but psychologists have argued instead that 'an underlying maladaptive personality disposition is conducive to the development of a worldview or worldviews that are more accepting of conspiracy theories'.[37] There is also evidence to suggest that conspiracy beliefs may fluctuate in response to context, emerging (and presumably also disappearing) as individuals make use of available cues in order to make sense of the world.[38]

Hypothesising that the often-observed negative association between education and conspiracy belief might be explained by lower prevalence of certain cognitive biases among more highly educated people, researchers have found belief in conspiracy theories to be predicted by anthropomorphism and the perception of intentionality where none exists;[39] perhaps relatedly, the relationship between education and conspiracy belief has been found to be mediated by belief in simple solutions and a perceived lack of control.[40] Both in the USA and in the Netherlands, individuals identifying as 'extremely left-wing' or 'extremely right-wing' have been found to be more likely to endorse conspiracy theories.[41]

32   Goertzel, Ted, "Belief in conspiracy theories." *Political Psychology* 15:4 (1994), pp.731–42.
33   Abalakina-Paap, Marina, and Walter G. Stephan, "Beliefs in conspiracies." *Political Psychology* 20:3 (1999), pp.637–47.
34   Swami, Viren, Rebecca Coles, Stefan Stieger, Jakob Pietschnig, Adrian Furnham, Sherry Rehim, and Martin Voracek, "Conspiracist ideation in Britain and Austria: evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories." *British Journal of Psychology* 102:3 (2011), pp.443–63.
35   See in particular Swami, Viren, Tomas Chamorro-Premuzic, and Adrian Furnham, "Unanswered questions: a preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs." *Applied Cognitive Psychology* 24 (2009), pp.749–61; Wood, Michael J., Karen M. Douglas, and Robbie M. Sutton, "Dead and alive: beliefs in contradictory conspiracy theories." *Social Psychological and Personality Science* 3:6 (2012), pp.767–73.
36   Goertzel, "Belief in conspiracy theories."
37   Swami, Viren, Laura Weis, Aixe Lay, David Barron, and Adrian Furnham, "Associations between belief in conspiracy theories and the maladaptive personality traits of the personality inventory for DSM-5." *Psychiatry Research* 236 (2016), pp.86–90.
38   Radnitz, Scott, and Patrick Underwood, "Is belief in conspiracy theories pathological? A survey experiment on the cognitive roots of extreme suspicion." *British Journal of Political Science* 47:1 (2017), pp.113–29.
39   Douglas, Karen M., Robbie M. Sutton, Mitchell J. Callan, Rael J. Dawtry, and Annelie J. Harvey, "Someone is pulling the strings: hypersensitive agency detection and belief in conspiracy theories." *Thinking & Reasoning* 22:1 (2016), pp.57–77.
40   van Prooijen, Jan-Willem, "Why education predicts decreased belief in conspiracy theories." *Applied Cognitive Psychology* 31 (2017), pp.50–58.
41   van Prooijen, Jan-Willem van, André P. M. Krouwel, and Thomas V. Pollet, "Political extremism predicts belief in conspiracy theories." *Social Psychological and Personality Science* 6:5 (2015), pp.578–579.

3 Conspiracy Theories and Violent Extremism

The history of right-wing violent extremism is inseparable from the history of conspiracism. Forged documents supposedly giving proof of a Jewish conspiracy were used to instigate pogroms, first in Tsarist Russia and then by the White forces after the Russian Revolution.[42] The ideas promoted within those documents were adopted by right-wing terrorist organisations in Weimar-era Germany and became central to the ideology of leading figures in the Nazi Party long before its rise to power.[43] Hitler, Goebbels and Rosenberg all doubted that the core document – the notorious *The Protocols of the Elders of Zion* – was genuine, but believed nonetheless that it expressed an 'inner truth'.[44] In pre-war Poland, nationalists mounted an attempted coup against the Mościcki government on the grounds that its leader was the agent of a supposed international Jewish-Masonic conspiracy; during the same period, the Japanese regime justified its attack on China on the grounds that the latter was controlled by 'Judeo-Masonry'.[45] One review of deadly terrorist attacks carried out on US soil by extreme right-wing individuals and groups from the 1980s onwards found that, in every case, there was *direct evidence* that the perpetrators believed in a Jewish conspiracy.[46] For example, Timothy McVeigh, the most lethal domestic terrorist in American history, was a long-term fan of author and radio host Milton Cooper, an internationally famous proponent of antisemitic conspiracy theories who was killed in 2001 after shooting a law enforcement officer in the head.[47]

Many more examples can be found. The Norwegian far-right terrorist Anders Breivik, who killed 77 people in 2011, was motivated by Islamophobic conspiracy theories.[48] Robert Bowers, who killed eleven worshippers at the Tree of Life synagogue in Pittsburgh, was motivated by belief in the 'great replacement', 'great substitution' or 'white genocide' conspiracy theory, which is traceable back to Adolf Hitler.[49] Manifestos endorsing the same conspiracy theory were released by Brenton Tarrant (who killed 51 worshippers at the Al Noor mosque in Christchurch), Patrick Crusius (accused of killing 22 at the Cielo Vista mall in El Paso), John Earnest (accused of killing one worshipper at the Chabad of Poway synagogue) and Philip Manshaus (accused of killing his younger sister and attempting to kill multiple worshippers at the

42   Cohn, *Warrant for genocide*, p.38, pp.118–19.
43   *ibid.*, pp.141–48, pp.179–82.
44   Bytwerk, Randall L., "Believing in 'inner truth': the Protocols of the Elders of Zion in Nazi propaganda, 1933–1945." *Holocaust and Genocide Studies* 29:2 (2015): pp.212–29.
45   Cohn, *Warrant for genocide*, pp.241–2.
46   Byington, Bradley, "Antisemitic conspiracy theories and violent extremism on the far right: a public health approach to counter-radicalization." *Journal of Contemporary Antisemitism* 2:1 (2019), p.1.
47   Byington, "Antisemitic conspiracy theories and violent extremism on the far right", pp.8–9.
48   Berntzen, Lars Erik, and Sveinung Sandberg, "The collective nature of lone wolf terrorism: Anders Behring Breivik and the anti-Islamic social movement." *Terrorism and Political Violence* 26 (2014): pp.759–79; Pantucci, Raffaello, "What have we learned about lone wolves from Anders Behring Breivik?" *Perspectives on Terrorism* 5:5-6 (2011): pp.27–42.
49   Allington and Joshi, "'What others dare not say'." pp.37–38.

Al-Noor Islamic Centre in Bærum).[50] Stephan Balliet, who killed two in an attempted mass shooting at a synagogue in Halle, denied the reality of the Holocaust and outlined an explicitly antisemitic version of the 'great replacement' conspiracy theory in a livestream begun just before the attack.[51] Before killing ten people at shisha bars in Hanau, Tobias Rathjen released a YouTube video announcing that the United States is 'under control of invisible secret societies' who 'abuse, torture, and kill little children in an unbelievable amount'. He exhorted all Americans to 'turn off the mainstream media' and 'fight now'.[52] None of these individuals appears to have been a member of an extremist organisation. In all cases, their radicalisation appears to have been the product of voluntary immersion in a largely online milieu saturated with conspiracy thinking.

Conspiracy theories also have a clear association with the extreme left. Popper presents Marx as a pioneering early critic of conspiracism, but laments that Marx's followers have 'put forward a popular conspiracy theory of society which is no better than the myth of the Learned Elders of Zion'.[53] Under Stalin, conspiracy theories were used to explain the 'incredible number of deaths in the first post-revolutionary decades, as well as the devastated state of the economy'.[54] The early 1930s saw accusations of involvement in an imaginary 'Trotskyist terrorist conspiracy',[55] and it became official Soviet doctrine that the USSR was threatened by an omnipresent network of 'imperialist' spies and saboteurs.[56] This culture continued beyond the death of Stalin: Soviet propaganda from the 1960s employed 'overtly antisemitic conspiracy theories and demonic portrayals of Jews and Zionists that echoed traditional European antisemitism'.[57] This propaganda was tremendously influential among ideologically sympathetic individuals and organisations, and it has been argued that, throughout 'the 1970s and the 1980s, the far-left in Britain and on the continent viewed Middle Eastern politics almost exclusively through the prism of Soviet anti-Zionism'.[58] Examination of declassified archive documents relating to the USSR's Propaganda Department suggests that this form of propaganda was motivated by beliefs that were sincerely held by senior members of the ruling Communist Party.[59]

Lastly, conspiracy theories have also been key to the motivation of religious extremism. Islamism is conspiracist at its heart: as Tibi observes, 'Islamists propagate the idea of a besieged Islam facing a *mu'amarah* (conspiracy) devised by *al-yahud wa al-salibiyun* (Jews and crusaders)'.[60] This worldview is the justification for the terrible crimes that jihadists commit against Muslims and non-Muslims alike.

50 Burke, Jason, "Norway mosque attack suspect 'inspired by Christchurch and El Paso shootings': online posts by Philip Manshaus praising other white extremist attacks emerge." *The Guardian* (11 August 2019); Noack, Rick, "Christchurch endures as extremist touchstone, as investigators probe suspected El Paso manifesto." *Washington Post* (6 August 2019); Darby, Luke, "How the 'Great Replacement' conspiracy theory has inspired white supremacist killers." *Daily Telegraph* (5 August 2019).
51 Chernick, Ilanit, Herb Keinon, and Benjamin Weintal, "Two killed in attack near synagogue in Halle, Germany. Gunman tried to blast way into shul, shot at passersby, kebab shop patrons. Rivlin calls on Germany to 'bring full force of law against antisemitism'." *Jerusalem Post* (10 October 2019), p.1.
52 Allington and Joshi, "'What others dare not say'." p.37.
53 Popper, *Conjectures and refutations*, p.125.
54 Yablokov, *Fortress Russia*, p.20.
55 Haslam, Jonathan, "Political opposition to Stalin and the origins of the terror in Russia, 1932–1936." *The Historical Journal* 29:2 (1986), p.395.
56 Shinar, Chaim, "Conspiracy narratives in Russian politics: from Stalin to Putin." *European Review* 26:4 (2018), pp.649–50.
57 Rich, Dave, *The left's Jewish problem: Jeremy Corbyn, Israel, and anti-Semitism*, 2nd ed. (London: Biteback Publishing, 2018), p.53.
58 Byford, *Conspiracy theories*, p.65.
59 Gjerde, Åsmund Borgen, "The logic of anti-Zionism: Soviet elites in the aftermath of the Six-Day War." *Patterns of Prejudice* 52:4 (2018), pp.271–92.
60 Tibi, Bassam, *Islamism and Islam* (New Haven: Yale University Press, 2012), p.58.

Moreover, it serves as a recruitment tool: in Islamist propaganda orientated towards audiences in the West, the 'Western media is framed as an extension of Crusader and Zionist interests, a propaganda tool mobilised by anti-Muslim forces in a war against Islam'.[61] Other forms of religiously motivated terrorism have also been linked to conspiracism, with perhaps the best example being Aum Shinrikyo, a syncretic Buddhist cult that has been designated a terrorist organisation in several countries. Two months before the most notorious of its many crimes – a chemical weapons attack on the Tokyo subway, which killed 13 and injured countless more – the cult 'published a crude antisemitic tract … in which it presented Japan's entire postwar history in terms of Jewish domination of the country' and 'formally declare[d] war on the "world shadow government"'.[62]

In addition to the above historical evidence, there is clear statistical support for the idea of an association not only between conspiracism and extremism, but also between conspiracism and the propagation of lawbreaking and ethnic or religious prejudice. Imhoff, Dieterle, and Lamberty (2021) find that conspiracy mentality is associated with reduced intention to engage in normative forms of political action such as voting, contacting politicians and taking part in legal demonstrations, and with increased intention to engage in non-normative forms of political action, such as vandalising property and committing violence against politicians and law enforcement officers.[63] On the basis of studies showing that the conspiracy mentality involves both perception of groups as powerful and hostility towards those same groups, Imhoff and Bruder argue that 'the mental shortcut of blaming individuals or groups may facilitate social action aimed at undermining the actions or goals of those perceived to be conspirators' but caution that 'social protest supported by conspiracy beliefs may also be particularly prone to turn ugly by targeting single groups or individuals and using them as scapegoats'.[64] Relatedly, experimental evidence suggests not only that exposure to conspiracy theories about immigrants increases anti-immigrant prejudice and that exposure to conspiracy theories about Jews decreases willingness to vote for Jewish political candidates, but also that exposure to conspiracy theories about Jews increases prejudice towards other outgroups.[65] Lastly, conspiracy beliefs have been found to predict criminal behaviour, and exposure to conspiracy theories has been found to increase intention to commit crime.[66]

---

61 Wiktorowicz, Quintan, *Radical Islam rising: Muslim extremism in the West* (Lanham, Maryland: Rowman and Littlefield, 2005), p.155.
62 Pipes, Daniel, *Conspiracy: how the paranoid style flourishes and where it comes from* (New York / London / Toronto / Sydney / Singapore: The Free Press, 1997), p.180.
63 Imhoff, Roland, Lea Dieterle, and Pia Lamberty, "Resolving the puzzle of conspiracy worldview and political activism: belief in secret plots decreases normative but increases nonnormative political engagement." *Social Psychological and Personality Science* 12:1 (2021), pp.71–9.
64 Imhoff, Roland, and Martin Bruder, "Speaking (un-)truth to power: conspiracy mentality as a generalised political attitude." *European Journal of Personality* 28 (2014), p.39.
65 Jolley, Daniel, Rose Meleady, and Karen M. Douglas, "Exposure to intergroup conspiracy theories promotes prejudice which spreads across groups." *British Journal of Psychology* 111 (2020), pp.17–35.
66 Jolley, Daniel, Karen M. Douglas, Ana C. Leite, and Tanya Schrader, "Belief in conspiracy theories and intentions to engage in everyday crime." *British Journal of Social Psychology* 58:3 (2019), pp.534–49.

4 Conspiracy Theories and Recent Political Violence in the United States

Recent years have seen conspiracy believers form extremist movements in their own right, creating amorphous milieus in which political violence is justified and even encouraged. A clear example is seen in 'Pizzagate'. This is the name given to a conspiracy theory which holds that children were abused and murdered in the basement of the Comet Ping Pong pizza restaurant by senior members of the US Democratic Party, allegations with clear echoes of medieval superstitions about Jews and witches (see section 1). The theory was promoted both by Alex Jones's InfoWars website and by the pro-Trump Breitbart News Network, as well as by a range of social media accounts and online celebrities.[67] On 1 December 2016, Edgar Maddison Welch encouraged his friends to join him in a 'raid' on Comet Ping Pong, telling them that they might have to 'sacrifice the lives of a few for the lives of many'.[68] On 4 December, Welch arrived at the restaurant armed with a knife, a handgun and an AR-15 assault rifle. After threatening staff, searching the premises and firing multiple shots, he was apprehended by police and subsequently sentenced to four years in prison.[69]

Further criminal acts were committed by Cesar Sayoc, who in 2019 was sentenced to 20 years' imprisonment for what the judge described as 'horrific acts of domestic terrorism'.[70] These acts involved constructing 16 pipe bombs and then mailing them to prominent Democrats and Democrat supporters. Delivered in October 2018, the crude devices would have been unlikely to explode, but their discovery led to the shutdown of many institutions, including schools.[71] According to his lawyers, Sayoc was '"connected" to hundreds of right-wing Facebook groups … [which] promoted various conspiracy theories' that he 'truly believed', and that served as a motivation for his crimes.[72] Sayoc's Twitter activity included the circulation of a large number of conspiracist memes, often alleging conspiracies involving some of the individuals he targeted.[73]

The 'QAnon' conspiracy theory holds that the USA is controlled by a cabal of paedophiles, against whom Donald Trump has secretly been waging war. As such, it recycles all the key elements of Pizzagate, while foregrounding its roots in medieval

67 Robb, Amanda, "Anatomy of a fake news scandal: inside the web of conspiracy theorists, Trump campaigners, and Twitter bots who manufactured the 'news' that Hillary Clinton ran a pizza-restaurant child-sex ring." *Rolling Stone* (16 November 2017).
68 *ibid.*
69 BBC, "'Pizzagate' gunman sentenced to four years." British Broadcasting Corporation (22 June 2017), https://www.bbc.co.uk/news/world-us-canada-40372407.
70 Ramey, Corinne, "Mail bomber Cesar Sayoc sentenced to 20 years in prison: Trump supporter had pleaded guilty to 65 felonies for mailing bombs to prominent Democrats across US." *Wall Street Journal* (5 August 2019), https://www.wsj.com/articles/mail-bomber-cesar-sayoc-sentenced-to-20-years-in-prison-11565040459.
71 United States v. Sayoc., United States District Court Southern District of New York.
72 Baumgartel, Sarah, Amy Gallicchio, and Ian Marcus Amelkin, "Re: United States v. Cesar Altieri Sayoc." (22 July 2019).
73 Allington, Daniel, and David Toube, "Conspiracy theories are not a harmless joke: alienated individuals are radicalised." *New Statesman* (16–22 November 2018), pp.15–16.

antisemitism and witch-hunting with explicit allegations of Satanism and the consumption of children's blood (modernised through pseudoscientific speculation about the alleged harvesting and use of adrenochrome). The theory arose from the collective interpretation of a series of highly cryptic messages, the first of which were posted to the 4chan message board in October 2017 by an anonymous user identified only as Q. It was given early support both by Alex Jones and by Breitbart's Curt Schilling.[74] While Trump has never claimed that the theory is true, he has given public support to its proponents since August 2018, when he invited Michael LeBron to the White House.[75] On 15 June 2018, Matthew Phillip Wright blocked a bridge over the Colorado River with an armoured vehicle, displaying handwritten signs calling upon Trump to release information referred to in messages from Q. After a short chase, Wright was apprehended by law enforcement officers, who found in his vehicle a handgun and an AR-15 assault rifle. He was sentenced to seven years and nine months in prison.[76] Further crimes apparently motivated by QAnon beliefs include a vehicle attack on the official residence of the Canadian prime minister and an alleged plan to detonate an explosive device in Springfield, Illinois.[77]

In May 2019, the FBI published a bulletin warning that conspiracy theories were very likely to motivate further criminal and violent activity, identifying the key moments of the 2020 election cycle as plausible flashpoints and specifically naming both Pizzagate and QAnon.[78] The FBI's concerns appeared vindicated by the storming of the Capitol building in Washington on 6 January 2021, which left five people dead, including a police officer (a sixth person, also a police officer, died by suicide shortly afterwards in an incident that one former official has linked to the riot; see McEvoy 2021).[79] This event took place following a demonstration at which Donald Trump had aired conspiracy theories alleging that widespread vote-rigging had cost him the US presidential election. After Trump directed protestors to march on the Capitol building, law enforcement officers were subjected to violent attacks with metal pipes and noxious substances, and parts of the building were overrun, with congresspeople, senators and the vice president having to be evacuated for their safety.[80] Offices were ransacked and computer equipment was stolen.[81] Suspected pipe bombs were left at the nearby headquarters of the Democratic and Republican parties,[82] and a police search uncovered eleven improvised firebombs, along with firearms and

74 Cassin, Elizabeth, and Mike Wendling, "What is the #QAnon conspiracy theory?". British Broadcasting Corporation (2018). Accessed 2 August: https://www.bbc.co.uk/news/blogs-trending-45040614.
75 Drury, Colin, "Trump meets 'paedophile cult' QAnon conspiracy theorist at White House." *The Independent* (25 August 2018), https://www.independent.co.uk/news/world/americas/qanon-trump-white-house-meeting-michael-lionel-lebron-conspiracy-theory-paedophile-ring-a8507766.html.
76 Associated Press, "Man gets prison for 2018 armed Hoover Dam bridge barricade." *Las Vegas Sun* (22 December 2020), https://lasvegassun.com/news/2020/dec/22/man-gets-prison-for-2018-armed-hoover-dam-bridge-b/.
77 Beckett, Lois, "QAnon: a timeline of violence linked to the conspiracy theory." *The Guardian* (16 October 2020), https://www.theguardian.com/us-news/2020/oct/15/qanon-violence-crimes-timeline.
78 FBI Phoenix Field Office, *(U//LES) Anti-government, identity based, and fringe political conspiracy theories*.
79 See McEvoy, Jemima, "Another Capitol police death: officer dies by suicide after responding to pro-Trump riot." *Forbes* (10 January 2021), https://www.forbes.com/sites/jemimamcevoy/2021/01/10/another-capitol-police-death-officer-dies-by-suicide-after-responding-to-pro-trump-riot/?sh=5488e9b570dd.
80 Tanfani, Joseph, John Shiffman, Brad Heath, Andrea Januta, and Mark Hosenball, "How security failures enabled Trump mob to storm U.S. Capitol." *Reuters* (blog) (7 January 2021), https://www.reuters.com/article/us-usa-election-capitol-security-insight-idUSKBN29C0R5.
81 Satter, Raphael, "U.S. senator says Capitol building rioters made off with laptop." *Reuters* (blog) (7 January 2021), https://www.reuters.com/article/us-usa-election-cyber/u-s-senator-says-capitol-building-rioters-made-off-with-laptop-idUSKBN29C2GA.
82 FBI, "Suspected pipe bombs in Washington, D.C." (6 January 2021), https://www.fbi.gov/wanted/seeking-info/suspected-pipe-bombs-in-washington-dc.

ammunition.[83] Rioters erected a symbolic gallows,[84] and a senior photojournalist repeatedly overheard them expressing the intention to kill the vice president;[85] video footage has emerged of a group of rioters chanting 'Hang Mike Pence'.[86]

Many of those who have since been accused of or identified as taking part in the insurrection were QAnon believers. For example, Roseanne Boyland, who died as a result of a medical emergency in the latter stages of the riot, was identified by her sister as a fervent QAnon believer who often shared false claims.[87] Ashli Babbitt, who was fatally shot while attempting to break into the Speaker's Lobby, had written the previous day that 'the storm' – the mythical denouement prophesied by Q – 'is here and it is descending upon DC in less than 24 hours'.[88] Jacob Chansley, or 'Jake Angeli', who is alleged to be the man photographed on the dais of the Senate carrying a spear,[89] was a prominent member of the QAnon believer community who referred to himself as the 'QAnon Shaman';[90] he was photographed with a handwritten sign declaring 'Q sent me'.[91] Further links are likely to emerge.

83 U.S. Attorney's Office, "Alabama man charged with possession of eleven Molotov cocktails found near protest at U.S. Capitol: authorities also found one assault rifle and three handguns." (8 January 2021), https://www.justice.gov/usao-dc/pr/alabama-man-charged-possession-eleven-molotov-cocktails-found-near-protest-us-capitol.
84 Godfrey, Elaine, "It was supposed to be so much worse: and the threat to the U.S. government hasn't passed." *The Atlantic* (9 January 2021).
85 Bourg, Jim, "I heard at least 3 different rioters at the Capitol say that they hoped to find Vice President Mike Pence and execute him by hanging him from a Capitol Hill tree as a traitor. It was a common line being repeated. Many more were just talking about how the VP should be executed." *@jimbourg* (blog) (8 January 2021), https://twitter.com/jimbourg/status/1347559078831284227.
86 Daily Kos, "Capitol insurrectionists chant 'Hang Mike Pence!', vow to return to finish job." (8 January 2021).
87 Thanawala, Sudhin, Stefanie Dazio, and Jeff Martin, "Family: Trump supporter who died followed QAnon conspiracy." *Associated Press News* (blog) (9 January 2021), https://apnews.com/article/election-2020-joe-biden-donald-trump-police-elections-7051411972c58cfbbf079876ce527ab4.
88 Trotta, Daniel, Gabriella Borter, and Jonathan Allen, "Woman killed in siege of U.S. Capitol was veteran who embraced conspiracy theories." *Reuters* (blog) (7 January 2021), https://www.reuters.com/article/us-usa-election-death/woman-killed-in-siege-of-u-s-capitol-was-veteran-who-embraced-conspiracy-theories-idUSKBN29C2NV.
89 U.S. Attorney's Office, "Three men charged in connection with events at U.S. Capitol." (9 January 2021), https://www.justice.gov/usao-dc/pr/three-men-charged-connection-events-us-capitol; United States Code Title 47 § 230. Protection for private blocking and screening of offensive material.
90 BBC, "'QAnon Shaman' Jake Angeli charged over pro-Trump riots." *BBC News* (blog) (10 January 2021).
91 Johnson, Jamie, "Jake Angeli: the 'QAnon shaman' from Arizona at the heart of the Capitol riots." *Daily Telegraph* (11 January 2021), https://www.telegraph.co.uk/news/2021/01/11/jake-angeli-qanon-shaman-arizona-heart-capitol-riots/.

5 Interventions Designed to Address the Digital Circulation of Conspiracy Theories

Empirical evidence has long existed that many individuals use digital communications media in order to 'spread falsehoods that strike them as plausible and that are consistent with their political predispositions'.[92] Accordingly, a notable research literature has grown up to address the question of what can be done to reduce this tendency.

Some findings are reassuring. There is evidence that simply encouraging people to reflect on conspiracy beliefs in an experimental context may reduce the negative effect of those beliefs,[93] and that prompting them to think about resistance to persuasion may reduce adherence to conspiracy beliefs themselves.[94] Moreover, while there are well-founded objections to political 'fact-checking' as it has often been practised in the press,[95] it seems that fact-checking can have a positive effect online: where social media reshares of a rumour receive responses that link to fact-checking web pages, those reshares appear more likely to be deleted,[96] and fact-checking labels have been found to diminish the persuasive effect of social media misinformation, especially if the labels are attributed to expert sources.[97]

On the other hand, there is also evidence that explicit corrections not only may fail to reduce pre-existing belief in false or unsubstantiated claims about political figures, but may, in some cases, even strengthen it, especially where individuals are strongly committed to belief in those claims or have ideological reasons for viewing them sympathetically.[98] Moreover, social media corrections are less likely to be accepted when the person offering the correction and the person being corrected have no pre-existing relationship,[99] and fact-checks appear to be most effective when the beliefs that they challenge are 'less engrained'.[100] There is also evidence that supportive comments on items of conspiracist content may have a tendency to receive more 'likes' than

92 Garrett, R. Kelly, "Troubling consequences of online political rumoring." *Human Communication Research* 37:2 (2011), p.270.

93 Einstein, Katherine Levine, and David M. Glick, "Do I think BLS data are BS? The consequences of conspiracy theories." *Political Behaviour* 37 (2014), pp.679–701.

94 Bonetto, Eric, Jaïs Troïan, Florent Varet, Grégory Lo Monaco, and Fabien Girandola, "Priming resistance to persuasion decreases adherence to conspiracy theories." *Social Influence* 13:3 (2018), pp.125–36.

95 Uscinski, Joseph E., and Ryden W. Butler, "The epistemology of fact checking." *Critical Review* 25:2 (2013), pp.162–80; Amazeen, Michelle A., "Revisiting the epistemology of fact-checking." *Critical Review* 27:1 (2015), pp.1–22; Uscinski, Joseph E., "The epistemology of fact-checking (is still naive): rejoinder to Amazeen." *Critical Review* 27:2 (2015), pp.243–52.

96 Friggeri, Adrien, Lada A. Adamic, Dean Eckles, and Justin Cheng, "Rumor cascades." Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (2014).

97 Zhang, Jingwen, Jieyu Ding Featherstone, Christopher Calabrese, and Magdalena Wojcieszak, "Effects of fact-checking social media vaccine misinformation on attitutdes toward vaccines." *Preventative Medicine* 145 (2021).

98 Nyhan, Brendan, and Jason Reifler, "When corrections fail: the persistence of political misperceptions." *Political Behaviour* 32 (2010), pp.303–30.

99 Margolin, Drew B., Aniko Hannak, and Ingmar Weber, "Political fact-checking on Twitter: when do corrections have an effect?" *Political Communication* 35:2 (2018), pp.196–219.

100 Lyons, Benjamin, Vittorio Merola, and Jason Reifler, "Not just asking questions: effects of implicit and explicit conspiracy information about vaccines and genetic modification." *Health Communication* 34:14 (2019), p.1748.

sceptical comments, hiding the latter from view.[101] All of this suggests that it will be very difficult for platforms to undo the damage done by the sharing of conspiracy theories, especially where those conspiracy theories have come to form part of an individual's or group's belief system and where communities defined by belief in conspiracy theories have come to provide a substantial proportion of an individual's social connections.

Much appears to depend on which messages reach an audience first. Anti-conspiracy arguments have been found to be effective if received before exposure to conspiracy theories, but not if received afterwards.[102] This supports the view that committed conspiracy believers cannot be reasoned with, and suggests the effectiveness of interventions based on inoculation theory, which holds that the persuasive effect of arguments is diminished by pre-exposure to refutations.[103] Encouragingly, a number of more recent studies have provided further evidence that such 'pre-bunking' of conspiracy theories can be effective.[104] However, this relies on being able to reach audiences before conspiracists do, particularly as the positive effect of inoculation has been found to be considerably reduced if audiences have been primed to reject inoculation with statements such as 'Next time some group or person "warns" you against listening to or thinking about an alternative message, ignore them'.[105]

The last twelve months have seen a massive increase in engagement with QAnon-related social media content.[106] However, Facebook did not act to ban QAnon-related groups until late in the year,[107] and when YouTube followed, it was with a more limited crackdown (more on which below). Questions also remain about how consistently platform policies are enforced with regard both to conspiracy theories and to other problematic forms of content. One investigation found that copies of QAnon videos removed from Facebook and Twitter remained in circulation on those same platforms,[108] and a cross-platform study of items of COVID-19 related content identified as misinformation by fact-checking organisations found that no action was taken with regard to 59% of such items on Twitter, 27% of such items on YouTube and 24% of such items on Facebook, despite platform policies which suggested that action would be taken promptly.[109] Moreover, analysis suggests that lack of coordination between platforms is further hampering the effectiveness of platform policies, with researchers observing that the 'strategy of reducing the visibility of misinformation

101 Allington and Joshi, "'What others dare not say'.", pp.35–53.
102 Jolley et al., "Belief in conspiracy theories.", pp.534–49.
103 Papageorgis, Demetrios, and William J. McGuire, "The generality of immunity to persuasion produced by pre-exposure to weakened counterarguments." *Journal of Abnormal and Social Psychology* 62 (1961), pp.475–81; for a meta-analysis that broadly confirms the general principle see Banas, John A., and Stephen A. Rains, "A meta-analysis of research on inoculation theory." *Communication Monographs* 77:3 (2010), pp.281–311.
104 Roozenbeek, Jon, Sander van der Linden, and Thomas Nygren, "Prebunking interventions based on 'inocultation' theory can reduce susceptibility to misinformation across cultures." *Harvard Kennedy School Misinformation Review* 1:2 (2020), pp.1–23; Maertens, Rakoen, Jon Roozenbeek, Melisa Basol, and Sander van der Linden, "Long-term effectiveness of inoculation against misinformation: three longitudinal experiments." *Journal of Experimental Psychology: Applied* advance online publication (2020).
105 Banas, John A., and Gregory Miller, "Inducing resistance to conspiracy theory propaganda: testing inoculation and metainoculation strategies." *Human Communication Research* 39 (2013), p.204.
106 Warzel, Charlie, "QAnon was a theory on a message board. Now it's headed to Congress." *New York Times*, (15 August 2020).
107 Luce, Edward, "QAnon cult shows America should fear the enemy within." *Financial Times* (11 September 2020), p.8.
108 Kaplan, Alex, "Facebook and Twitter said they removed a viral video spreading false claims about COVID-19. Copies are still circulating." *Media Matters* (blog) (28 July 2020), https://www.mediamatters.org/coronavirus-covid-19/facebook-and-twitter-said-they-removed-viral-video-spreading-false-claims.
109 Brennen, J. Scott, Felix Simon, Philip N. Howard, and Rasmus Kleis Nielsen, *Types, sources, and claims of COVID-19 misinformation.* Reuters Institute for the Study of Journalism (Oxford) (7 April 2020), https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation; CCDH, *#WilltoAct: how social media giants have failed to live up to their claims on the Coronavirus 'infodemic'* (London: Center for Countering Digital Hate, 2020).

on one platform will not be successful if the same content is shared tens of thousands of times on another platform'.[110]

In a similar vein, investigators found that QAnon-linked creators banned from YouTube were able to continue broadcasting via Apple TV, Roku and Amazon Fire,[111] and to receive thousands of dollars per month via supporters on Patreon.[112] Moreover, Facebook's strategy of featuring news from trustworthy sources has given Breitbart a role as a provider,[113] when (as we have seen) the latter has played a central role in the dissemination of conspiracy theories linked to violent extremism. YouTube's policy with regard to content featuring conspiracist claims that have been 'used to justify real-world violence' involves removing such content only on condition that it 'targets an individual or group', which provides a loophole through which much dangerous content is likely able to pass.[114]

The evidence presented in this section suggests that, in order to prevent conspiracy beliefs from taking hold, it will be necessary to engage individuals with rational argument *before* they are exposed to conspiracy theories, and at a stage where their social connections have not yet come to be dominated by conspiracy believers. This is not compatible with facilitating the dissemination of conspiracy theories and the formation of groups devoted to conspiracy thinking. Truly addressing the problem might therefore require a fundamental platform redesign. Researchers have argued that, as it is people with an existing tendency towards conspiracy thinking who are most likely to be swayed into new conspiracy beliefs, it is those people who need the most protection from conspiracy materials online.[115] However, the outcome that leading social networking and media sharing sites were designed to achieve is almost the exact reverse: Facebook brings people with similar interests together; YouTube recommends videos that are deemed similar to those to which the viewer has already responded positively; and Twitter and Instagram recommend users to follow accounts similar to those they are already following. Such functionality carries an inherent risk because of its obvious potential to lead susceptible individuals into progressively more extreme views.

---

110 Knuutila, Aleksi, Aliaksandr Herasimenka, Hubert Au, Jonathan Bright, Rasmus Nielsen, and Philip N. Howard, *Covid-related misinformation on YouTube: the spread of misinformation videos on social media and the effectiveness of platform policies* (Oxford: Oxford Internet Institute, 2020).

111 Kaplan, Alex, and Olivia Little, "Apple TV, Roku, and Amazon Fire are allowing QAnon channels on their streaming platforms - mostly thanks to one video hosting service." *Media Matters* (blog) (20 November 2020), https://www.mediamatters.org/qanon-conspiracy-theory/apple-tv-roku-and-amazon-fire-are-allowing-qanon-channels-their-streaming.

112 Little, Olivia, "Patreon is profiting from QAnon." *Media Matters* (blog) (20 October 2020), https://www.mediamatters.org/qanon-conspiracy-theory/patreon-profiting-qanon.

113 Darcy, Oliver, "Facebook News launches with Breitbart as a source." *CNN* (blog) (26 October 2019), https://edition.cnn.com/2019/10/26/media/facebook-news-breitbart/index.html.

114 Murphy, Hannah, and Siddarth Venkataramakrishnan, "YouTube joins Twitter in crackdown on pro-Trump conspiracy theory QAnon." *Financial Times* (16 October 2020), p.1.

115 Klofstad, Casey A., Joseph E. Uscinski, Jennifer M. Connolly, and Jonathan P. West, "What drives people to believe in Zika conspiracy theories?" *Palgrave Communications* 5 (2019).

6 Conclusion: a Cultural Change for Platforms

Much of the conventional wisdom that shapes the internet of today was formed in an earlier online world where the need for many checks and balances had not yet become apparent. Before the commercialisation of the web, internet users were a small minority disproportionately comprised of individuals associated with academic institutions, research organisations, and tech companies. As such, it could function as a free-for-all in which activists could credibly maintain the fiction that governments 'have no sovereignty' and real-world 'legal concepts … do not apply'.[116] Such ideas were soon established to be false, but assumptions formed in relation to the pre-commercial internet continued to guide platform policy even when internet access had become mainstream.[117] The gulf between the worldview of the early tech entrepreneurs and that of wider society was exposed by the Unite the Right rally at Charlottesville:

> *When several tech companies kicked alt-right users off their platforms after Charlottesville, they were met with a vigorous backlash from many in the industry. Matthew Prince, CEO and co-founder of Cloudflare, who reluctantly banned virulently racist site,* The Daily Stormer*, from his service … fretted about the decision. 'As [an] internet user, I think it's pretty dangerous if my moral, political, or economic whims play some role in deciding who can and cannot be online,' he said. The Electronic Frontier Foundation issued a statement that read, in part, 'we believe that no one … should decide who gets to speak and who doesn't.'[118]*

It is tempting to explain the position adopted in the above both by the Cloudflare CEO and by the Electronic Frontier Foundation with reference to the First Amendment to the United States Constitution (1791). However, the First Amendment acts only to limit the power of government and does nothing to limit a citizen or corporation's right to engage in the kind of decision-making that the business of publishing has always involved. The idea that the First Amendment guarantees all speakers the right to a platform has no legal reality – nor any relationship to the basis on which mainstream platforms are actually run. Such platforms typically impose blanket bans on certain forms of content, such as pornography, and engage in active content moderation, especially by deleting posts and by adding warnings. Moreover, while the United States legislation popularly referred to as 'Section 230' is often invoked as a counter-argument to this view, the famous declaration that '[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider' must be understood in its legal context as one of a series of '[p]rotection[s]

116   Barlow, John Perry, *A declaration of the independence of cyberspace*. Electronic Frontier Foundation (1996).
117   Cohen-Almagor, Raphael, "Freedom of expression, Internet responsibility, and business ethics: the Yahoo! saga and its aftermath." *Journal of Business Ethics* 106:3 (2011), pp.353–65.
118   Daniels, Jessie, "The algorithmic rise of the 'alt-right'." *Contexts* 17:1 (2018), p.62.

for … blocking and screening of offensive material'.[119] Crucially, these also include the declaration that '[n]o provider or user of an interactive computer service shall be held liable on account of … any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be … objectionable, whether or not such material is constitutionally protected'.[120] In other words, if Facebook chose to close all conspiracy theory groups or if YouTube chose to delete all conspiracist content regardless of whether it targeted individuals, those decisions would be explicitly protected under US law. As one legal scholar put it, 'Section 230 was designed to free online forums to police bad content without becoming legally liable for all that they missed. But many early tech-company lawyers missed the lesson'.[121]

Throughout the first decade of the 21st century, the natural relationship between the tech industry and the cultural industries was widely supposed to be one in which the latter would produce content for free distribution through a neutral infrastructure provided by the former.[122] That world is now gone, in part because the commercial consequences for producers so often turned out to be dire.[123] For several years now, the most promising commercial developments have been around companies that have persuaded customers to pay for a high quality product. Amazon's business, for example, has always been built around selling products for a fee, which is why it has been able to invest in award-winning content for its subscription video streaming service.[124] At Netflix, the leader in the streaming market, expenditure on content has risen not only in absolute terms, but also relative to its number of subscribers.[125] Even YouTube, at one time a website closely associated with uncontrolled sharing of copyright-infringing content, now offers two premium services and a paid television service with a total of more than 22 million subscribers.[126] The tremendously successful launch of the Disney+ streaming service last year led to predictions of vast growth;[127] by the end of the fiscal year, it was already exceeding expectations in terms of numbers of subscribers.[128]

The world thus appears already to have moved away from a situation in which the model internet company was one that made money by allowing its users to produce or share content while selling a share of their attention to its advertisers, and towards a situation in which internet companies compete for revenue by providing customers with goods, services and experiences that those customers consider valuable. As investors adapt to this new commercial reality, social networking and media sharing companies founded in the first decade of the century will need to ask themselves *why* they should continue to be associated with forms of content that history, experimental research

---

119   United States Code Title 47 § 230. Protection for private blocking and screening of offensive material.
120   *ibid.*
121   Franks, Mary Anne, Mike Godwin, Jess Kosseff, and Andrés Martinez, "Where do we go from here with Section 230? Three legal scholars discuss the interner law that everyone seems to hate right now." *Slate* (15 December 2020).
122   See, in particular, Anderson, Chris, *Free: the future of a radical price* (London: Random House, 2009).
123   For a perceptive early critique of the 'free' model, see Levine, Robert, *Free ride: how the internet is destroying the culture business, and how the culture business can fight back* (London: The Bodley Head, 2011).
124   Csathy, Peter, "Amazon Prime Video: the stealthy, ominous streaming force." *Forbes* (31 January 2020).
125   Trefis Team, "Neflix, one question: is it losing money or making money?" *Forbes* (1 May 2020).
126   Statt, Nick, "YouTube is a $15 billion-a-year business, Google reveals for the first time: we've never before known how much money YouTube generates." *The Verge* (3 February 2020).
127   Del Vecchio, Gene, "Disney Plus has the potential to become a $30 billion giant in only five years." *Forbes* (11 May 2020).
128   Business Wire, *The Walt Disney company reports fourth quarter and full year earnings for fiscal 2020* (12 November 2020), https://www.businesswire.com/news/home/20201112006058/en/.

and recent political events have associated with a host of negative outcomes. In a world where quality content sells, why would anyone want his or her company to be known as one of the leading distributors of misinformation, even if that misinformation takes a seemingly harmless form, such as flat-earthism or moon-landing denial?

Throughout much of the twentieth century, the major publishers and broadcasters performed an invaluable service by ensuring that the toxic blend of fantasy and superstition that Michael Barkun calls 'stigmatized knowledge' remained subordinated to real knowledge in the public sphere.[129] Since then, conspiracy theory and pseudoscience have been leaking from the margins into the mainstream at an ever-increasing rate, thanks less to the technical affordances of the internet than to its custodians' misguided belief that to deny falsehoods a mass audience would be morally wrong. The view that content moderation is intrinsically unethical, or that it ceases to be so only once the content in question has crossed some sort of threshold, is an atavism – a holdover from the internet's elitist past – of benefit to no one but mountebanks and demagogues. We ought to be past that now.

Embracing the changes that the internet has undergone since the 1980s will mean moving beyond an exclusive focus on negative measures when considering how to mitigate the problem that conspiracism undoubtedly represents. It is doubtful that many people sign up for social media accounts with the intention of exposing themselves to depressing and enraging untruths. As well as acting to remove misinformation, platforms can legitimately make a positive choice to disseminate information from reputable sources and to provide users with tools by which they may seek out life-affirming online experiences and interactions. It is hard to imagine serious objections to such a move. Wikipedia, for example, has always recognised and prioritised the authority of traditional sources of knowledge,[130] which is one reason why it has not descended into a post-truth morass and can credibly be treated as a source of factual information by both Alexa and Siri, as well as by Google and Facebook.[131] Social networking and media sharing platforms would do well to learn from its approach.

129 Barkun, *A Culture of Conspiracy*.
130 Tagg, Caroline, "Digital English." In *Communicating in English: talk, text, technology*, edited by Daniel Allington and Barbara Mayor (Abingdon: Routledge, 2012), p.319.
131 The Economist, "The other tech giant." *The Economist* (9 January 2021), pp.56–7.

Policy Landscape

This section is authored by Armida van Rij and Lucy Thomas, both Research Associates at the Policy Institute based at King's College London. It provides an overview of the relevant policy landscape for this report.

Introduction

The spread of misinformation, including disinformation, reaches the core of democratic societies. It poses a threat to democratic processes, undermines policymaking outcomes and further increases deeply entrenched polarisation across political beliefs and communities. Yet it remains a particularly difficult area to tackle. Should responsibility lie with technology companies or should governments regulate more keenly? Can regulation sit alongside freedom of speech and the right to free thinking? These are the kind of challenging questions with which policymakers, technology companies and civil society organisations grapple.

Twitter's recent permanent suspension of the account of US President Donald Trump following the Capital Hill riot, for which commentators widely believe Trump has some responsibility due to his repeated rejections of the legitimate US election result, has thrown a spotlight on these policy challenges once again. In this report, we aim to give an overview of what nine jurisdictions are doing in terms of tackling the spread of misinformation.

Conspiracy Theories, Radicalisation and Digital Media: Assessing The Policy Developments and Challenges

Canada

The Canadian government's counter-radicalism strategy encompasses traditional intelligence and security agency activities, engagement with civil society, collaborative initiatives with industry, and community-focused policing. Its National Strategy on Countering Radicalization to Violence has three core areas of action: to engage with civil society, to support countering violent extremism (CVE) research and to collaborate with international initiatives and tech companies.[132]

Canada places a strong emphasis on counter-messaging and engagement with civil society. Extreme Dialogue is a counter-messaging initiative between the Canadian government and the Institute for Strategic Dialogue. The project provides

---

132 "National Strategy on Countering Radicalization to Violence," Public Safety Canada. Accessed: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ntnl-strtg-cntrng-rdclztn-vlnc/index-en.aspx#s7.

educational resources to practitioners and young people through films that illustrate the negative impact of extremism.[133] The Canada Centre for Community Engagement and Prevention of Violence coordinates a number of community-based interventions to counter radicalisation. In Calgary, for instance, the ReDirect programme works with the Calgary Police Service and the City of Calgary Community & Neighborhood Services, as well as health and social services agencies to intervene in the early stages of radicalisation. The programme employs a range of strategies including referral, education and providing advice for individuals seeking a way to leave a violent extremist group.[134]

Canada has not been immune from the spread of conspiracy theories. In the winter of 2020, there were a number of anti-mask and COVID-19-denier rallies in major cities, which included the presence of Canadian QAnon supporters.[135] A QAnon researcher and commentator contends that 'One of the biggest QAnon promoters on social media is Canadian'.[136] This community-based and education-focused strategy extends to Canada's efforts to counter the spread of dangerous conspiracy theory content. The government invested $7 million in 2019 and 2020 in a robust digital literacy campaign for citizens in order to 'strengthen citizens' critical thinking about online disinformation [and] their ability to be more resilient against online disinformation.'[137] A further $3.5 million has been invested to counter COVID-19 misinformation online. As a final part of this Digital Citizen Initiative, Canada is also funding a multi-stakeholder strategy of research and engagement in order to 'build citizen resilience against online disinformation and building partnerships to support a healthy information ecosystem.'[138]

European Union

For the EU, disinformation became a priority security issue following Russia's annexation of Crimea in 2014, during which Russia began to use disinformation extensively as part of its armoury in conducting hybrid warfare. An East StratCom Task Force was established to monitor and understand how Russia's disinformation infiltrated into Western media outlets and general discourse. The task force has built a database of over 8,000 examples of disinformation on its website EUvsDisinfo.eu and has developed extensive media monitoring and strategic communications capabilities, although this project is due to finish at the end of 2021.[139]

Despite tackling disinformation remaining a priority, the EU still grapples with the lack of consensus on basic issues regarding disinformation. For example, among member states, many do not

133 See: https://extremedialogue.org/.
134 See: http://redirect.cpsevents.ca/.
135 Montpetit, J. and J. Macfarlane, "Anti-mask protest in Montreal draws large crowd, propelled by US conspiracy theories," *CBC* (12 September 2020). Accessed: https://www.cbc.ca/news/canada/montreal/anti-mask-protest-montreal-1.5722033.
136 Kovac, A., "How Canada became one of the world's biggest hubs for QAnon conspiracy theories," *CTV News* (3 November 2020). Accessed: https://montreal.ctvnews.ca/how-canada-became-one-of-the-world-s-biggest-hubs-for-qanon-conspiracy-theories-1.5172097.
137 "Online disinformation," Government of Canada. Accessed: https://www.canada.ca/en/canadian-heritage/services/online-disinformation.html
138 *ibid.*
139 Pamment, J., "The EU's Role in Fighting Disinformation: Taking Back the Initiative," Carnegie Endowment (15 July 2020). Accessed: https://carnegieendowment.org/2020/07/15/eu-s-role-in-fighting-disinformation-taking-back-initiative-pub-82286.

recognise disinformation as problem per se, or do not 'publicly attribute particular malign activities to the offending adversaries'.[140] Within the EU institutions, lack of coordination and ownership stalls progress.[141]

In 2018, the European Commission introduced a self-regulatory Code of Practice on Disinformation for social media platforms, advertisers and online platforms to 'address the spread of online disinformation and fake news'.[142] This included examples of best practice, which includes principles such as 'platforms endeavour to tackle disinformation by pursuing follow the money approaches to disinformation and preventing bad actors from receiving remuneration'.[143]

The code has had mixed results and opinions differ over its level of success. The EC carried out a targeted monitoring of the implementation of the commitments made by the signatories in the first half of 2019. It found that the 'Code has proven a very valuable instrument' in terms of being a framework for structured dialogue between stakeholders. However, both the assessment itself and critics of the code have found shortcomings. Critics believe the code does not go far enough in addressing disinformation.[144]

In late 2018, the EC launched the Action Plan Against Disinformation, which places disinformation in the context of hybrid threats. The Action Plan has four key goals: 1) improve detection, analysis and exposure of disinformation; 2) stronger cooperation and joint responses to disinformation; 3) mobilise private sector to tackle disinformation; 4) raise awareness and improve societal resilience.[145] Under the plan, the EC established the Rapid Alert System (RAS), which facilitates cooperation with international partners. The RAS has strengthened 'cooperation with online platforms, to identify and prevent the spread of disinformation campaigns'.[146] As with the Code of Practice on Disinformation, while its intentions may be worthwhile, the use of the RAS remains limited. There is currently limitied information-sharing and engagement. One positive outcome, however, has been the ability for small coalitions of likeminded states to form and act jointly.[147]

The Code of Practice and the Action Plan are some of the few policy levers the EC has developed in recent years. Others include the convening of the High-Level Expert Group on Fake News and Online Disinformation and efforts to ensure the 2019 European Parliamentary elections would not be disrupted by disinformation campaigns.

140 *ibid.*
141 *ibid.*
142 "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions – Tackling Online Disinformation: A European Approach," European Commission (26 April 2018). Accessed: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236.
143 "Annex II Current Best Practices from Signatories of the Code of Practice," European Commission. Accessed: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54455.
144 Stolton, S., "EU code of practice on disinformation 'insufficient and unsuitable,' member states say," *EURACTIVE* (5 June 2020). Accessed: https://www.euractiv.com/section/digital/news/eu-code-of-practice-on-disinformation-insufficient-and-unsuitable-member-states-say/.
145 "Action Plan Against Disinformation: Report on progress," European Commission (June 2019). Accessed: https://ec.europa.eu/commission/sites/beta-political/files/factsheet_disinfo_elex_140619_final.pdf.
146 *ibid.*
147 Pamment, "The EU's Role in Fighting Disinformation".

France

France approved law number 2018-1202 on the 'fight against the manipulation of information' in 2018. It aims to 'better protect democracy against the different ways in which fake news is deliberately spread'.[148] The focus in France lies in particular on politically sensitive times around elections. During these periods, the law stipulates that first, there is 'a transparency obligation for digital platforms, who need to report any sponsored content by publishing the name of the authors and the amount paid. Platforms exceeding a certain number of hits a day must have a legal representative in France and publish their algorithms.'[149] Second, the law creates 'a legal injunction allowing the circulation of fake news to be swiftly halted'.[150] Outside election periods, the law assigns a 'duty of cooperation' to social media companies and online platforms to tackle fake news.[151] The French Broadcasting Authority (CSA) has been tasked with ensuring compliance with these measures. It also has the authority to 'prevent, suspend and stop the broadcasts of television services that are controlled by foreign states'.[152] The law has been subject to significant criticism, with opponents arguing that it stifles free speech.[153]

Prior to the law coming into effect, France had a range of other legal tools at its disposal. The 1881 French Press Law prohibits the 'publication, distribution, or reproduction by whatever means of "false news" or "articles fabricated, falsified or falsely attributed to others" where this is done in bad faith and undermines, or could undermine, public order'.[154] This law was effectively updated with the June 2004 law on online communications.

In the aftermath of the Capitol Hill riot of January 2021, German Chancellor Angela Merkel called for more regulation on online incitement, rather than leaving the regulation of free speech up to social media companies and online platforms.[155] This call was echoed by French Finance Minister Bruno Le Maire.[156]

Ghana

Governmental strategic efforts to combat violent extremism online in Ghana are limited, since political violence in the country has not been fuelled by terrorist activities. The Global Terrorism Database, a database of global terror attacks since 1970, lists only 21 incidents with 23 fatalities in 50 years in Ghana.[157] Ghana therefore has a strong civil society presence to commission research, engage citizenry, lobby the government and coordinate action on pressing

---

148 Government of France, "Against information manipulation." Accessed: https://www.gouvernement.fr/en/against-information-manipulation.
149 *ibid.*
150 *ibid.*
151 *ibid.*
152 *ibid.*
153 Fiorentino, M., "France passes controversial 'fake news' law," *Euronews* (22 November 2018). Accessed: https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law.
154 Craufurd Smith, R., "Fake news, French Law and democratic legitimacy: lessons for the United Kingdom," *Journal of Media Law*, 11:1 (2019), p.55. Accessed: https://www.tandfonline.com/doi/pdf/10.1080/17577632.2019.1679424?needAccess=true&.
155 "Angela Merkel attacks Twitter over Trump ban," *Financial Times*. Accessed: https://www.ft.com/content/6146b352-6b40-48ef-b10b-a34ad585b91a.
156 Bruno Le Maire: 'Je suis convaincu que dans le deuxième temps de 2021, l'économie française rebondira,' France Inter (11 January 2021). Accessed: https://www.franceinter.fr/emissions/l-invite-de-8h20-le-grand-entretien/l-invite-de-8h20-le-grand-entretien-11-janvier-2021.
157 Global Terrorism Database, START. Accessed: https://www.start.umd.edu/gtd/.

societal matters, including countering misinformation and conspiracy theories online.

In a 2019 survey, 69% of respondents had encountered misinformation over the past year.[158] One of the biggest challenges Ghana faces in terms of its digital ecosystem is the dominance of well-funded political parties on social media in the run-up to its elections in December 2020. Civil society groups voiced concern that 'wealthy politicians' … social media machines drown out the voices of smaller parties.'[159] During the coronavirus pandemic, Ghana struggled to contain a viral video, which many claimed to be President Nana Akufo-Addo, that included dangerous claims about the origins of the virus, that the pandemic was a planned event and that the government was making vaccines mandatory.[160]

GhanaFact, a project launched by FactSpace West Africa, was established as an independent non-profit social enterprise in 2019 to counter disinformation online. GhanaFact established a presence on Twitter and Facebook during the December 2020 elections in order to actively fact-check dangerous misinformation and conspiracy theories about military disruption of voting, power blackouts and burning ballot boxes.[161] Due to inadequate resources, however, non-profit organisations have a limited reach and effectiveness.

In keeping with other regions of the world, Africa has engaged in multi-stakeholder discussions to counter disinformation and conspiracy theories relating to the pandemic. The African Telecommunications Union coordinated with the International Telecommunications Union and the UN Under Secretary-General/ Special Advisor working on Digital Cooperation in April 2019 to discuss how to manage the so-called 'infodemic'. In Kenya for instance, one such outcome was to extend internet access to traditionally limited areas by building connectivity infrastructure.[162]

In December 2020, the World Health Organisation launched the Africa Infodemic Response Alliance (AIRA) in order to coordinate regional responses to the disinformation crisis. AIRA 'brings together 13 international and regional organizations and fact-checking groups with expertise in data and behavioural science, epidemiology, research, digital health, and communications to detect, disrupt and counter damaging misinformation on public health issues in Africa'. It will also support national efforts to combat harmful content online through robust research, recruiting specialists and deploying engagement strategies for credible informational sources.[163]

158 "About," GhanaFact. Accessed: https://ghanafact.com/about/.
159 Bax, P. and Prinsloo, L., "Online Disinformation Campaigns Undermine African Elections." *Bloomberg Businessweek* (13 October 2020). Accessed: https://www.bloomberg.com/news/articles/2020-10-13/disinformation-campaigns-on-facebook-twitter-google-undermine-african-election.
160 Goodman, J. et al., "Coronavirus: Fact-checking fake stories in Africa," *BBC* (8 August 2020). Accessed: https://www.bbc.co.uk/news/53684037.
161 Oduor, M., "GhanaFact: Combating misinformation during election," *AfricaNews* (23 December 2020). Accessed: https://www.africanews.com/2020/12/22/ghanafact-combating-misinformation-during-elections/.
162 "Countering Disinformation and Misinformation in Global Response to COVID-19," NewsGhana (1 May 2020). Accessed: https://newsghana.com.gh/countering-disinformation-and-misinformation-in-global-response-to-covid-19/.
163 "Landmark alliance launches in Africa to fight COVID-19 misinformation," *The Herald* (8 December 2020). Accessed: https://www.herald.co.zw/landmark-alliance-launches-in-africa-to-fight-covid-19-misinformation/.

Japan

Due to its ageing population, Japan has conventionally been understood to rely on traditional media, such as newspapers and television, to a large extent. A 2016 survey found that Japanese citizens spent the least amount of time on social media in the world.[164]

However, as younger generations turn to social media more, online disinformation and conspiracy theories have spread in Japan. So-called 'summary sites' (まとめサイト), which aggregate information from around the internet, particularly commentary and opinion pieces rather than traditional fact-checked reporting, are incredibly prevalent on Japanese social media, according to one analysis. The analysis found that the most shared online article about President-elect Joe Biden in 2020 was one shared by a summary site named 'Anonymous Post', that claimed that the US National Guard had been deployed to counter voter fraud and that the voting rate in Wisconsin was 200%. The article had over 23,000 shares on Twitter and Facebook.[165]

The aggregation of unverified disinformation, originating in the USA, on Japanese summary sites has fuelled the spread of conspiracy theories and may have contributed to pockets of support for Donald Trump in Japan. In late November 2020, hundreds of people in Tokyo marched in support of Trump following his loss in the early November general election.[166] Reporting has found that popular Japanese accounts on Twitter coordinate to spread pro-Trump and QAnon content online.[167]

Analysts of the growing Trump support movement in Japan point to wider social fissures and changes as reasons for its emergence and spread. Yasushi Watanabe, a professor at Keio University, contends that traditional societal norms, such as 'the notion of Japanese being homogenous', are breaking down and high-profile scandals implicating senior governmental officials have contributed to a context where criticism of '*jokyu kokumin*' ('privileged citizens') has emerged. In this context of profound social change, anxiety and fear, conspiracy theories and a politics based on fear, such as Trumpism, are likely to gain traction.

In late 2019, the Japanese government signalled its intent to partner with leading global social media networks – Google, Facebook, Apple and Amazon – to combat disinformation online.[168] However, the government is moving too slowly on the topic of digital literacy and education in order to foster a healthy media ecosystem online in Japan. The consequences of such inaction could prove to be devastating, as we have seen play out in many tragic instances around the globe.

164 "Japanese spend least time on social media worldwide," *Nikkei Asia* (18 August 2016). Accessed: https://asia.nikkei.com/Business/Japanese-spend-least-time-on-social-media-worldwide.
165 Kota Hatachi, "Information about Mr Biden's 'illegal' win has spread in large numbers in Japan. Are summary sites or new religious media influential?" *Buzzfeed Japan* (13 November 2020). Accessed: https://www.buzzfeed.com/jp/kotahatachi/election2020.
166 Kuni Miyake, "Who were those Trumpists marching in Tokyo?" *The Japan Times* (15 December 2020). Accessed: https://www.japantimes.co.jp/opinion/2020/12/15/commentary/japan-commentary/trump-protests-tokyo/.
167 St. Michel, P., "US election helps disinformation on social media thrive in Japan," *The Japan Times* (21 November 2020). Accessed: https://www.japantimes.co.jp/news/2020/11/21/national/media-national/u-s-election-fake-news-social-media/.
168 "Japanese panel wants to establish team to fight fake news, with help from US tech giants," *The Japan Times* (30 November 2019). Accessed: https://www.japantimes.co.jp/news/2019/11/30/business/japan-fake-news-gafa/.

New Zealand

New Zealand's overarching counter-terrorism strategy, released in February 2020, involves the coordination of multiple agencies and bodies to counter harmful content online.[169] Similar to Canada (above), these bodies range from the Cabinet External Relations and Security Committee, to police, intelligence and security communications agencies, to foreign affairs, trade, defence, transport, innovation and development agencies.

The Christchurch mosque shootings in March 2019 showed that New Zealand is also impacted by conspiracy theory content online. Brenton Tarrant, the perpetrator of the attacks, released a manifesto on an under-regulated message board site based on the 'Great Replacement' conspiracy theory, which contends that white Western men are being threatened by immigration and 'feminising' phenomena, such as greater visibility of trans people.

To combat such conspiracy-driven content online, New Zealand has taken up a leadership position globally in championing cross-country and cross-sector initiatives. Most notably, in the aftermath of the Christchurch shootings, the governments of New Zealand and France brought together a coalition of heads of state with social media and technology companies under the Christchurch Call to Eliminate Terrorist and Violence Extremist Content Online.[170] Signatories to the call are committed to enforce laws that prohibit the dissemination of terrorist and violent extremist content online, yet also to respect freedom of expression and privacy concerns. The countries also work to support capacity-building and awareness-raising activities in order to prevent the use of online services to disseminate terrorist and violent extremist content.

The Christchurch Call also commits companies, including Amazon, Facebook, Google, Twitter, Facebook and YouTube, to greater industry standards of accountability and transparency. The companies must enforce their community standards and terms of service by prioritising content moderation and removal actions, and identifying content in real-time for review and assessment. Collectively, the countries and companies are developing efforts with civil society to promote community-led activities in order to intervene in the processes of online radicalisation.

The Christchurch Call also acted as the vehicle through which the Global Internet Forum to Counter Terrorism (GIFCT) was overhauled. As part of the overhaul, GIFCT's remit expanded to include a suite of preventative, response and educational activities in the effort to counter violent extremism and disinformation online.[171] New Zealand's efforts to co-sponsor a range of cross-sector global initiatives showcase a more horizontal approach to governing extremists' use of tech platforms. The approach encompasses conventional security and intelligence structures as well as initiatives that bring together practitioners, academia, policymakers and tech leaders to formulate responses to harmful online content.

169 Government of New Zealand, Officials' Committee for Domestic and External Security Coordination, Counter-Terrorism Coordination Committee, "Countering terrorism and violent extremism national strategy overview" (February 2020). Accessed: https://dpmc.govt.nz/sites/default/files/2020-02/2019-20 CT Strategy-all-final.pdf.
170 See: https://www.christchurchcall.com/.
171 Global Internet Forum to Counter Terrorism, "Next Steps for GIFCT" (23 September 2019). Accessed: https://gifct.org/press/next-steps-gifct/.

United Kingdom

The United Kingdom's approach to combatting extremist use of online platforms follows a traditional mode of governance in which state institutions take the lead. The central institution responsible for counter-terrorism legislation is the Home Office, which also coordinates with the Government Communications Headquarters, the country's security and intelligence organisation. The Home Office has also created collaborative bodies with other government institutions (most often the Department for Digital, Culture, Media, and Sport) and Parliament, such as the UK Council for Internet Safety, the National Counter Terrorism Security Office and the Commission on Countering Extremism.[172]

The central activity of the UK's approach to countering disinformation online is around the regulation of social media and technology platforms. The government's Online Harms White Paper, published in April 2019, set out a comprehensive case for greater national regulation of social media.[173] Under this new regulatory framework, social media and technology companies will bear a new statutory duty of care to their users, enforceable via Ofcom, the UK's regulatory body for communications. Ofcom will subject platforms to financial and technical penalties – websites could be blocked at ISP level and fined up to 4% of their global turnover – for non-compliance with the framework and violations of the statutory duty of care.[174] At the time of writing, the Online Harms Bill, the legislative operationalisation of the White Paper, has been delayed for several years.[175]

In January 2018, then UK government announced the establishment of a National Security Communications Unit in order to clamp down on the spread of fake news and disinformation.[176] The announcement came amid a political climate in which the government was under pressure to investigate claims of Russian bots and internet farms skewing discourse online with regard to the 2016 Brexit referendum.[177] However, no further details emerged regarding the National Security Communications Unit and no search results are available on the UK's government website, suggesting that the project has been shuttered.

As QAnon and COVID-19 conspiracy theories continue to gain traction in the UK, the government urgently needs to coordinate a strategic and robust response to countering the spread of dangerous misinformation online.

172 Gov.uk, UK Council for Internet Safety. Accessed: https://www.gov.uk/government/organisations/uk-council-for-internet-safety; Gov.uk, Commission for Countering Extremism. Accessed: https://www.gov.uk/government/organisations/commission-for-countering-extremism; Gov.uk, National Counter Terrorism Security Office. Accessed: https://www.gov.uk/government/organisations/national-counter-terrorism-security-office.
173 HM Government, 'Online Harms White Paper' (April 2019). Accessed: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.
174 Crawford, A., "Online Harms bill: Warning over 'unacceptable' delay," *BBC* (29 June 2020). Accessed: https://www.bbc.co.uk/news/technology-53222665.
175 *ibid.*
176 "Government announces anti-fake news unit," *BBC*, (29 January 2018). Accessed: https://www.bbc.com/news/uk-politics-42791218.
177 "Disinformation and 'fake news': Interim Report: Fifth Report of Session 2017–19," House of Commons Digital, Culture, Media and Sport Committee. Accessed: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf.

UN Counter-Terrorism Executive Directive

The UN Counter-Terrorism Committee Executive Directorate (UN CTED) was established by UN Security Council Resolution 1535 (2004) as an expert body in support of the Security Council's Counter-Terrorism Committee (CTC).[178] Its aim was to assess UN Member States' implementation of Security Council resolutions on counterterrorism, and support their efforts through dialogue. The UN CTED works closely with the Security Council, the private sector – in particular, social media companies and online platform providers – and civil society organisations.

UN CTED is concerned about the use of disinformation for political end-goals. In its trend alert of July 2020, it writes: 'Member States and researchers have warned that extreme right-wing terrorists are using Covid-19-related conspiracy theories and disinformation to radicalise, recruit and fundraise, as well as seeking to inspire plots and attacks.'[179] The concern is that the fear caused by COVID-19 has proved to be a fertile recruitment ground for right-wing terrorist organisations, for which conspiracy theories may act as 'radicalisation multipliers'.[180]

In an attempt to tackle coronavirus-related disinformation, the UN has launched the Share Verified Initiative in collaboration with other organisations. The purpose of the initiative is to share trusted information about COVID-19 and to appeal to communities to share fact-based information and thereby counter COVID-19 disinformation.[181] This is alongside private sector initiatives and Member State initiatives.

United States

The United States' policy approach to combatting the misuse of tech platforms during the Trump administration has been damaging both nationally and internationally. Although the USA has co-sponsored various cross-national initiatives, such as Tech Against Terrorism and the Global Counterterrorism Forum, the Trump administration has undermined any good faith efforts to combat the spread of conspiracy theories online.

In terms of historical national policy, the Department of Homeland Security (DHS), the Department of Justice, the Federal Bureau of Investigation, the National Counter Terrorism Center and the National Security Council and Congress, among others, have been at the forefront of the response.[182] 'Counter messaging, awareness briefings, partnerships, and legislation' are all methods that have been piloted.[183]

178 Chowdhury Fink, N., "Meeting the challenge: A guide to United Nations counterterrorism activities," International Peace Institute (2012), p.45. Accessed: https://www.ipinst.org/wp-content/uploads/publications/ebook_guide_to_un_counterterrorism.pdf.

179 United Nations Security Council Counter-Terrorism Committee Executive Directorate, "Member States Concerned by the Growing and Increasingly Translational Threat of Extreme Right-Wing Terrorism" (July 2020), p.1. Accessed: https://www.un.org/sc/ctc/wp-content/uploads/2020/07/CTED_Trends_Alert_Extreme_Right-Wing_Terrorism_JULY.pdf.

180 *ibid.*

181 United Nations Department of Global Communications, "'Verified' initiative aims to floor digital space with facts amid COVID-19 crisis" (28 May 2020). Accessed: https://www.un.org/en/coronavirus/%E2%80%98verified%E2%80%99-initiative-aims-flood-digital-space-facts-amid-covid-19-crisis.

182 Alexander, A., "A Plan for Preventing and Countering Terrorist and Violent Extremist Exploitation of Information and Communications Technology in America," George Washington University Program on Extremism (2019), p.5. Accessed: https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/A%20Plan%20for%20Preventing%20and%20Countering%20Terrorist%20and%20Violent%20Extremist.pdf.

183 *ibid.*

In 2011, the Obama administration established the Countering Violent Extremism Task Force in order to 'unify the domestic CVE effort'.[184] The task force was intended to bring together practitioners from the bodies listed above in order to coordinate engagement with civil society, develop intervention models, create investments in research and cultivate communications and digital strategies.[185] However, in early 2017, Trump considered restructuring the task force to remove white supremacist terrorism from its remit, renaming the programme the 'Countering Radical Islamic Extremism'.[186] Furthermore, a budget unveiled in the spring of 2017 cut all funding to countering violent extremism programmes.[187] By late October 2018, the task force had shuttered, as funding expired and 'staff members returned to their home agencies and departments'.[188]

Trump's actions reveal a deep hostility towards CVE efforts generally, but specifically those aimed at community outreach and engagement with local civil society and those targeting far-right and white supremacist terrorism. For instance, one of the recipients of DHS funding was Life After Hate, an initiative that works with individuals to leave white supremacist and neo-Nazi groups.[189] Removing funding and curtailing the remit to exclude white supremacy from the USA's efforts demonstrated the Trump administration's implicit support for white supremacist and racist terrorist actions.

Should the incoming Biden administration fail to establish a robust strategy to counter dangerous online content, the global consequences could be deadly. As we have seen with the 6 January 2021 insurrection at the Capitol, and as armed militias prepare to attack at the presidential inauguration, as QAnon and anti-mask rallies spread across the country and across the globe, the unchecked spread of conspiracy theories has terrifying and tragic real-world implications.

Moderating conspiracy theory content in the United States: ethical questions and challenges

On 8 August 2018, YouTube, Facebook, Apple and Spotify removed online content by the far-right conspiracy theorist and radio talk show host Alex Jones, including clips from his call-in radio show InfoWars.[190] The rationale for the removal was that Jones had breached the companies' terms of service on hate speech, by claiming that European nations were in danger of 'being taken over by Muslim immigrants'.[191] The following month, Twitter permanently suspended Jones and his InfoWars accounts, after Jones claimed that former President Barack Obama was the 'global head of Al-Qaeda',

184 Department of Homeland Security, "Countering Violent Extremism Task Force." Accessed: https://www.dhs.gov/cve/task-force.
185 *ibid.*
186 Ainsley, J. et al., "Exclusive: Trump to focus counter-extremism program solely on Islam – sources," *Reuters* (3 February 2017). Accessed: https://www.reuters.com/article/idUSKBN15G5VO?feedType=RSS&feedName=topNews&utm_source=twitter&utm_medium=Social.
187 Ainsley, J., "White House budget slashes 'countering violent extremism' grants," *Reuters* (23 May 2017). Accessed: https://www.reuters.com/article/us-usa-budget-extremism-idUSKBN18J2HJ.
188 Beinart, P., "Trump Shut Programs to Counter Violent Extremism," *The Atlantic* (29 October 2018). Accessed: https://www.theatlantic.com/ideas/archive/2018/10/trump-shut-countering-violent-extremism-program/574237/.
189 Life After Hate, "About Us." Accessed: https://www.lifeafterhate.org/about-us-page.
190 Hern, A., "Facebook, Apple, YouTube and Spotify ban Infowars' Alex Jones," *The Guardian* (6 August 2018). Accessed: https://www.theguardian.com/technology/2018/aug/06/apple-removes-podcasts-infowars-alex-jones.
191 "YouTube removes 'hate speech' videos from InfoWars," *BBC* (26 July 2018). Accessed: https://www.bbc.com/news/technology-44965160.

and that 'transgenderism [is an] evil paedophile plot to sexualise and destroy children.'[192]

However, Jones has repeatedly found ways to appear on social media platforms, for example Facebook livestreaming from a newly created 'Infowars is Back' page in May 2019,[193] and by being hosted on the massively popular *The Joe Rogan Show* podcast in October 2020.[194] At the time of writing, Jones still operates an InfoWars storefront on Amazon, from which he sells vitamin supplements and survivalist products,[195] which at one time included a colloidal silver toothpaste that he claimed would cure coronavirus.[196] Content from the InfoWars website and from the daily talk show can still be posted on Facebook and Twitter. I myself have been able to share an out-of-context humorous InfoWars clip with friends and colleagues on social media.[197] The bans also drove Jones to seek out new platforms in order to grow his following; for example, his Instagram following grew 57% over seven months after his Facebook ban.[198]

Banning, moderating or removing conspiracy theory content online is fraught with social, legal and ethical challenges. There are no absolute or easy solution to the problem. This analysis will take a look at how conspiracy theories function in the cultural discourse, the effects of the commodification of conspiracy theorising and how conspiracy theories moved from being counter-cultural to highly visible in today's discourse. The piece goes on to reflect on three ethical challenges of moderating conspiracy theory content online: whether social media companies should act as gatekeepers, whether moderation of conspiracy theorising causes harm and how to approach the moderation of conspiracy-theory-as-entertainment content.

Conspiracy theories can be defined as a type of theorising about past and present events in the world that is based on a number of overlapping concepts, namely: actions and events are marked by intentionality and by secrecy and deception; coincidence and structural explanations are impossible, and instead events are defined by causality and correlation; explanations for events are highly detailed and complex; and explanations for events are based on a blend of fact and fiction.[199] However, in considering the challenges in moderating or banning conspiracy theory content online, perhaps the most important aspect of conspiracy theories is their productive nature.

By productive nature, we refer to the ways in which conspiracy theories produce identities and therefore function as important social, political and cultural drivers. Conspiracy theorising 'code[s] and express[es] actual socio-cultural and political concerns and anxieties',[200] such as

192 *ibid.*
193 Haselton, T., "Alex Jones was banned from Facebook, but an hour later he was back on Facebook livestreaming," *CNBC* (2 May 2019). Accessed: https://www.cnbc.com/2019/05/02/alex-jones-banned-from-facebook-but-hes-already-back.html.
194 Waterson, J., "Joe Rogan hosts Alex Jones on Spotify podcast despite ban," *The Guardian* (28 October 2020). Accessed: https://www.theguardian.com/technology/2020/oct/28/joe-rogan-hosts-alex-jones-on-spotify-podcast-despite-ban.
195 See: https://www.amazon.com/stores/Infowars+Life/page/5358EFBF-442C-467C-AFCD-983AA4F4C233?ref_=ast_bln.
196 Higgins-Dunn, N., "NY attorney general orders InfoWars' Alex Jones to stop selling coronavirus 'treatment' products," *CNBC* (12 March 2020). Accessed: https://www.cnbc.com/2020/03/12/ny-attorney-general-orders-alex-jones-to-stop-selling-coronavirus-cures.html.
197 See: https://twitter.com/rightwingwatch/status/1202347123125182464.
198 Martinez, N., "Instagram is the new home for Alex Jones and Infowars," *Media Matters* (19 March 2019). Accessed: https://www.mediamatters.org/alex-jones/instagram-new-home-alex-jones-and-infowars.
199 Barkun, *A Culture of Conspiracy*, pp.3–4.
200 Thalmann, K., *The Stigmatization of Conspiracy Theory since the 1950s: 'A Plot to Make us Look Foolish,'* (London: Routledge, 2019), p.3.

in the cases of the people who conducted attacks in the name of the 'Great Replacement' theory investigated in the report above, whose anxieties about a relative decline in the white population found both a delusional explanation and a violent expression. Conspiracy theorising produces strong identities of both the self – 'patriots' or 'rebels' – and the conspirators – evil beings 'plotting the self's enslavement or destruction'.[201] In this way, the cultural work of conspiracy produces collective identities based on enmity, evil, struggle and salvation, as well as profoundly defining 'who we are'.

For this reason, conspiracy theories have operated throughout time to self-define within elite discourse. For instance, the Red Scare of 1950s America, in which leading members of the political, social and spiritual establishment believed there was a concerted Communist effort to undermine and ultimately destroy the nation, was a firmly accepted part of the mainstream and elite-sponsored discourse. Over time, as Katharina Thalmann in *The Stigmatization of Conspiracy Theory since the 1950s* demonstrates, the epistemological model of conspiracy theorising became increasingly marginalised and stigmatised.[202] Academic theories on conspiracy theory, building on Karl Popper's conceptualisation, explained conspiracy theorising by individual pseudoscience – for example, Richard Hofstadter's seminal work, *The Paranoid Style in American Politics* – as 'a means to defend the sciences at a time when the hunt for communist subversives affected university departments all over the country'.[203] Over time, conspiracy theorising fell along firm epistemological boundaries of inclusion and exclusion, pushing conspiracy theorising out of accepted mainstream discourse and into the 'paranoid' fringes of society.

We might therefore accept Michael Barkun's theory that conspiracy theories can be understood as 'stigmatized knowledge … [which is] knowledge claims that have been ignored or rejected by those institutions we rely upon to validate such claims.'[204] Universities, the media, religious authorities and the medical community are all examples of 'institutions [that] provide forms of implied or direct "certification" that ideas, beliefs, or fact assertions can be relied upon.'[205] Conspiracy theories therefore play a role in the struggle to define what is legitimate and what is illegitimate knowledge in the cultural discourse, a discourse that 'negotiate[s] what is sayable and unsayable, filter[s] the legitimate from the illegitimate.'[206] This boundary is not a stable one, but rather fluid and unfixed, constantly being negotiated and shifting through time and place.

If we consider conspiracy theories as that which produces social and cultural identities and that which is widely understood to be illegitimate, stigmatised knowledge, we can begin to understand the emergence of what Thalmann calls 'a veritable counter-discourse on conspiracy theory and conspiracist counter-culture'.[207] This has led to the proliferation of 'superconspiracy theories', in which multiple conspiracies become nested in one another to create a complex, sprawling network of conspiratorial actors across the globe.

201 Butter, M., *Plots, Designs, and Schemes: American Conspiracy Theories from the Puritans to the Present* (Berlin: Freiberg Institute for Advanced Studies, 2014), p.21.
202 Thalmann, *The Stigmatization of Conspiracy Theory*.
203 *ibid.*, p.14.
204 Barkun, M., 'Conspiracy Theories as Stigmatized Knowledge', *Diogenes* (2016). Accessed: https://doi.org/10.1177/0392192116669288.
205 *ibid.*
206 Thalmann, *The Stigmatization of Conspiracy Theory*, p.13.
207 *ibid.*, p.12.

Superconspiracy theories were popularised in the late 20th century by 'celebrity' conspiracy theorists, such as David Icke, and accelerated in reach with the advent of the internet and social media in the past three decades, such as we see with Alex Jones's InfoWars and the rise of QAnon.

Indeed, the continual marginalisation and stigmatisation of conspiracist claims is a vital aspect of the success of these conspiracy theories. The counter-culture thrives on its outsider status, tapping into the anxieties of those who also define themselves by their outsider or critical thinker subject position. Indeed, the exclusion of conspiracy theory from mainstream cultural discourse functions as evidence for continued complicity by elites and hegemonic institutions.

The rise of popular culture that features conspiracy theories – television programmes such as *The X-Files* and *Stranger Things*, international bestselling books such as Dan Brown's *The Da Vinci Code* – in the late 20th century points to the process by which conspiracy theorising has become popularised and entered the mainstream.[208] The effect is a widespread socialisation into conspiracy theory culture: through popular culture, audiences have become familiarised with the 'style and rhetoric, tropes and devices recurrent in conspiracy theories', so that conspiracy theories have become increasingly visible in the public sphere and continue to retain their appeal 'as a conceptual model' to approach events in the world.[209] Conspiracy theories are, as Clare Birchall argues, 'now part of our collective response to local and global events'.[210]

As cultural discourse is increasingly played out on the internet and social media, the commodification of conspiracy theorising has similarly been taking place online. The 2010s saw the rise of online provocateurs and conspiracy theorists, such as Alex Jones, Donald Trump and Milo Yiannopolous, who literally capitalised from the peddling of conspiracy theories and by re-emphasising their ostensible outsider status.[211] Thalmann contends that by 'posing as a relentless arbiter of truth in opposition to traditional media (Jones), as an outsider to Washingtonian politics challenge a conspiratorial and corrupt elite (Trump), and as a right-wing subcultural jester (Yiannopolous), all three have been able to convert conspiracy theory's stigma into social and economic capital.'[212]

Conspiracy theorising, then, has inched out of the realms of counter-culture and is visible once more in widespread cultural discourse over the past decade or so. As the report above demonstrates, social media plays an incontrovertible role in the manufacture, dissemination and consumption of conspiracy theories. The sheer monetary success of 'professional' conspiracy theorists online such as Jones suggests the presence of a vast conspiracy theory industry, where there is considerable profit to be made.

208 Birchall, C., *Knowledge Goes Pop: From Conspiracy Theory to Gossip* (Oxford: Berg, 2006), p.38.
209 Thalmann, *The Stigmatization of Conspiracy Theory*, p.191.
210 Birchall, *Knowledge Goes Pop*, p.34
211 Court records show that, by 2014, Alex Jones' businesses were netting $20 million a year in revenue, and that he bought four Rolex watches in one day in 2014. See: Williamson, E. and Steel, E., "Conspiracy Theories Made Alex Jones Very Rich. They May Bring Him Down," *New York Times* (7 September 2018). Accessed: https://www.nytimes.com/2018/09/07/us/politics/alex-jones-business-infowars-conspiracy.html.
212 Thalmann, *The Stigmatization of Conspiracy*, p.197.

With this framework in place, we can begin to think through the implications and challenges of aggressive moderation and banning of conspiracy theory content online.

Should Social Media Corporations Act as Gatekeepers of Conspiracy Theorising?

Over the years, social media companies have been tasked with policing harmful content on their platforms. As we have seen above, several platforms moved to remove InfoWars content in mid-2018. In 2019, Instagram and Facebook removed a number of far-right conspiracy theorists, including Yiannopoulos, Paul Joseph Watson and Laura Loomer, from its platforms.[213] In August 2020, Facebook expanded its policy to include 'violence-inducing conspiracy networks' such as QAnon.[214] Twitter permanently suspended Donald Trump in early January 2021.

Although these people have produced, distributed and profited from conspiracy theories, their removal from social media has often been for 'comorbid' issues such as breaching terms of service for incitement to violence, hate speech or harassment. In the end, these platforms do not moderate conspiracy theorising online that does not directly incite or induce violence.

Jack Dorsey's rationalisation of initially not banning Jones from Twitter is instructive here: 'We didn't suspend Alex Jones or Infowars yesterday. We know that's hard for many but the reason is simple: he hasn't violated our rules. We'll enforce if he does … Accounts like Jones' can often sensationalize issues and spread unsubstantiated rumours, so it's critical journalists document, validate, and refute such information directly so people can form their own opinions. This is what serves the public conversation best.'[215] A liberal interpretation of the First Amendment means that social media CEOs can often treat conspiracy theories as simply another opinion in the marketplace of ideas.

This opens up a number of tricky ethical issues. First, it positions corporations as gatekeepers of the social and cultural discourse. Ultimately, social media platforms are corporations that operate solely for profit. Making the argument that such individuals and companies should wield such power over discourse, public health and public safety – or indeed that they have the public interest at heart – is a difficult one.

Secondly, as we have seen above, conspiracy theories can be understood as stigmatised knowledge. That means that any knowledge dismissed by institutions could conceivably be labelled as a conspiracy theory. For instance, distrust in the medical establishment by black people in the United States is prevalent,[216] due to continued

---

213 "Instagram and Facebook Ban Far-Right Extremists," *The Atlantic* (May 2018). Accessed: https://www.theatlantic.com/technology/archive/2019/05/instagram-and-facebook-ban-far-right-extremists/588607/.

214 Facebook, "An Update to How We Address Movements and Organizations Tied to Violence" (19 August 2020). Accessed: https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/.

215 See: https://twitter.com/jack/status/1026984249960755200.

216 Armstrong, K. et al., "Racial/Ethnic Differences in Physician Distrust in the United States," *American Journal of Public Health*, 97:7 (2007). Accessed: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1913079/; Williams, C., "Black Americans don't trust our healthcare system – here's why," *The Hill* (24 August 2017). Accessed: https://thehill.com/blogs/pundits-blog/healthcare/347780-black-americans-dont-have-trust-in-our-healthcare-system.

slavery-era assumptions by medical professionals that black people do not feel pain[217] and unethical medical experiments such as the Tuskegee syphilis experiment.[218] Only recently have the academic, scientific and media establishments begun to affirm this reality; prior to doing so, this knowledge would have been invalidated as conspiracy theorising. As social media becomes the new gatekeepers of discourse, the question becomes whether conspiracy theories should be curtailed, and if so, how that could be done without imperilling users' free speech.

Lastly and relatedly, moderating conspiracy theorising online can only ethically be undertaken at the point at which it spills over into other harm, such as hate speech, harassment and even inciting violence. This approach is one that is often 'too little, too late', as we have seen with the QAnon phenomenon. Failing to quash the theory adequately and, indeed, pushing users towards QAnon content means that by the time platforms take aggressive action it is too late.[219] The mass banning of Trump and other accounts that spread conspiracy theories came only after the 6 January 2021 riot at the US Capitol and a fractious general election in which a QAnon supporter was elected to Congress,[220] two examples of the devastating real-world consequences of inaction or delayed action.

Does Policing Conspiracy Theories Online Cause More Harm?

Moderating or removing conspiracy theorists from major social media platforms has two unintended and arguably equally dangerous consequences. As we have seen above, in order to protect certain forms of stigmatised knowledge, social media companies can only moderate conspiracy theories online ethically once real-world harm has been threatened or already committed. By the time Jones was kicked off social media, the damage was already done: he was already globally famous, his content had enjoyed years of exposure and reach, and he had been financially enriched. This social and economic capital means that Jones continues to exact and to contribute to real-world harms. For example, Jones personally financed $500,000 towards booking The Ellipse, a park between the White House and the Washington Monument in Washington D.C., for the rally on 6 January 2021 that led to the insurrection at the US Capitol.[221] Additionally, Jones led an anti-mask rally outside the Texas state capitol, endangering public health at the height of the coronavirus pandemic.[222]

217 Vergano, D., "Many Doctors Hold Racist Beliefs About How Black People Feel Pain," *Buzzfeed News* (27 April 2016). Accessed: https://www.buzzfeednews.com/article/danvergano/race-pains.

218 "40 Years of Human Experimentation in America: The Tuskegee Study," McGill Office for Science and Society. Accessed: https://www.mcgill.ca/oss/article/history/40-years-human-experimentation-america-tuskegee-study.

219 "Facebook has also acknowledged that pages and groups associated with QAnon extremism had at least 3 million members, meaning Facebook helped radicalize 2 million people." See: McNamee, R., "Platforms Must Pay for Their Role in the Insurrection," *Wired*, (7 January 2021). Accessed: https://www.wired.com/story/opinion-platforms-must-pay-for-their-role-in-the-insurrection/.

220 Levin, S., "QAnon supporter Marjorie Taylor Greene wins seat in US House," *The Guardian* (4 November 2020). Accessed: https://www.theguardian.com/us-news/2020/nov/03/qanon-marjorie-taylor-greene-wins-congress.

221 Graziosi, G., "Alex Jones says he paid $500,000 for rally that led to Capitol riot," *The Independent* (11 January 2021). Accessed: https://www.independent.co.uk/news/world/americas/us-election-2020/alex-jones-paid-rally-capitol-riot-b1784603.html.

222 Palmer, E., "Infowars' Alex Jones Leads 'We Can't Breathe' Anti-Mask Protest in Texas," *Newsweek* (29 June 2020). Accessed: https://www.newsweek.com/texas-covid-alex-jones-mask-protest-1514068.

The too-little-too-late moderation of conspiracy theorising online can therefore translate into martyrdom, whereby 'professional' conspiracy theorisers are enabled to carry out more extreme actions, perhaps even violence, offline. Since followers of conspiracy theories already lean into their 'outsider' status, being deplatformed or banned can push adherents further down the path of radicalisation .

Relatedly, when conspiracy theorists are banned from or limited on social media platforms, this can entrench belief as further evidence of the conspiracy. It could even expand the scope of the conspiracy theory to include the media; believers will take 'the "debunking" or blocking of conspiracist content as evidence of the bias exhibited by the mainstream media or even as a sign that the media, too, are part of an elite conspiracy'.[223]

Time will tell whether Donald Trump, who flirted with and enabled white power militia groups up until the 6 January 2021 insurrection, will migrate to another platform to continue to spread conspiracy theories, or whether the ban will contribute to a dampening of the violent potential of conspiracy theories. The point is that when platforms act in isolation from one another or take action without shouldering the consequences that will play out on another platform, we enter dangerous territory. Past experience of blocking and banning Islamic State and al-Qaeda content online has shown that it creates a supply-and-demand opportunity for un- and under-regulated platforms to fill the void.[224] Mainstream platforms must create transparent contingency plans in collaboration with other platforms and internet service providers when they take aggressive action against figures as powerful as Trump and Jones.

Should Conspiracy Theories Shared for Entertainment be Moderated?

In late January 2019, YouTuber Shane Dawson premiered a documentary mini-series, *Conspiracy Series with Shane Dawson*, to his 20 million subscribers.[225] The first episode, 'Conspiracy Theories with Shane Dawson', is an hour and forty-four minutes long and opens with a 50-second sponsorship advert for an online coupon service. The video blends vlog-style content with scenes of Dawson driving his fiancé and friends through areas affected by the 2018 California wildfires. As they drive, Dawson asks questions, such as 'How does every house on the street catch fire except one?' On another street, houses on one side are totally destroyed while those on the other side are unaffected. 'What does that *mean*?' Dawson asks rhetorically. 'I'm not a scientist… but that's, like, something.' Other topics he explores in the video include children being manipulated by dark themes in cartoons, clothing brands using subliminal messaging and iPhones monitoring their users.[226] Although Dawson does not explicitly endorse a belief in any conspiracy theory, his viewers are presented with 'evidence'

223 Thalmann, *The Stigmatization of Conspiracy Theory*, p.195.
224 See, for example, a previous GNET report about alternatives to the encrypted messaging platform, Telegram: https://gnet-research.org/wp-content/uploads/2020/11/GNET-Report-Migration-Moments-Extremist-Adoption-of-Text%E2%80%91Based-Instant-Messaging-Applications_V2.pdf.
225 See: https://socialblade.com/youtube/user/shane/monthly.
226 See: https://youtu.be/BHLBaOASC74.

that leans towards accepting a conspiratorial explanation for the events – which, arguably, has the same deleterious effect as explicitly conspiratorial content.

Initially, the video was demonetised after the automatic moderation service identified within it a harmful prank that had been banned by the site.[227] Dawson wrote on Twitter that he thought his videos weren't 'brand friendly' but that 'advertisers should get on board w/ edgier stuff.'[228] However, the video was remonetised shortly thereafter, with a YouTube spokesperson describing the initial action as an error.[229] At the time of writing, the video has amassed over 49.5 million views, earning Dawson hundreds of thousands of dollars in advertising revenue.[230]

The week prior to the premiere of 'Conspiracy Theories with Shane Dawson', YouTube had released a statement announcing a pilot policy on conspiracy theory content. The company said that 'We'll begin reducing recommendations of borderline content and content that could misinform users in harmful ways – like videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11.' Although these videos would not be featured on YouTube's algorithm where they would be recommended content to most users, they could still be recommended to YouTube users who are subscribed to a channel uploading this 'borderline' content.[231]

YouTube's policy and its support for Dawson's conspiracy series raise difficult ethical and practical questions around moderating conspiracy content that is intended to be entertainment. In presenting conspiracy theories as possibilities, Dawson's series and other conspiracy-adjacent content challenges YouTube in its quest to balance freedom of speech with responsible moderation. YouTube has long been under fire for the platform's approach to recommending similar content, which can to radicalisation, putting increasingly radical ideas in front of users.[232]

Dawson's videos are ultimately entertainment, but they also socialise viewers into becoming comfortable with the patterns of thinking common to conspiracy theorising. For instance, his continual rhetorical questioning of 'what does this *mean*?' affirms that there is a secret, possibly sinister meaning of and explanation for events in the world. His huge influence on the platform normalises conspiracy theories, allowing his viewers to be more comfortable talking about conspiracy theories.

It isn't possible to parse out the true impact of conspiracy-as-entertainment content online, but a reasonable conclusion is that it contributes to a normalisation of conspiracy theorising.

227 Alexander, J., "YouTube's vague conspiracy theory policies present issues for the platform," *The Verge* (6 February 2019). Accessed: https://www.theverge.com/2019/2/6/18213648/shane-dawson-conspiracy-video-youtube-demoneitzation-community-guidelines.
228 See: https://twitter.com/shanedawson/status/1091161227076657152.
229 Alexander, J., "YouTube's reviewing mistake could cost Shane Dawson thousands of dollars" *The Verge* (31 January 2019). Accessed: https://www.theverge.com/2019/1/31/18205059/shane-dawson-conspiracy-theories-youtube-demonetized
230 Based on the average that ad revenue of $2 per 1,000 views, inflated for Dawson's reach and popularity. See: Alexander, "YouTube's reviewing mistake".
231 Newton, C., "YouTube says it will recommend fewer videos about conspiracy theories," *The Verge* (29 January 2019). Accessed: https://www.theverge.com/2019/1/25/18197301/youtube-algorithm-conspiracy-theories-misinformation.
232 Tufekci, Z., "YouTube, The Great Radicalizer," *New York Times* (10 March 2018). Accessed: https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

This does not necessarily need to mean that it is a first step along a radicalisation path and that such content should therefore be moderated aggressively, nor that it contributes to a general and total re-legitimisation of conspiracy theorising in contemporary discourse. It does, however, contribute to the general heightened visibility of conspiracy theories online, which has the potential to bolster the beliefs of conspiracy theorists. More worryingly, it allows content creators online to capitalise on conspiracy theories, opening the door to grifters such as Alex Jones who have amassed a small personal fortune by pushing harmful misinformation and inciting real-world violence.

Global Network
on Extremism & Technology