MEMORANDUM OF UNDERSTANDING

TO: RT HON NADINE DORRIES MP, MINISTER for DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT

FROM: A N Other

DATE: 22nd September 2021

RE: **Legal Opinion on the Best Possible Means and Digital Tools for the Department for Digital, Culture, Media, & Sport to Review and Implement in Order to Successfully Tackle the Spread of Conspiracy Theories on Social Media**

Combating the Spread of Conspiracy Theories on Social Media

Conspiracy theories – we've all come across them, and, according to statistics, the majority of the population will have believed in at least one in their life.[1] But what makes up a conspiracy theory, who are the people that believe in them, why are they so problematic, and why does social media's existence seem to have amplified their spread? In the past 18 months, COVID-19 and the events of January 6th at The Capitol in Washington DC has meant that conspiracy theories and the threat they pose to society has come to the fore.

This paper aims to address the problems posed by the proliferation of conspiracy theories across Western mainstream social media platforms – particularly focusing on Facebook, Twitter, Instagram, and YouTube. Whilst conspiracy theories have existed in a myriad of ways across humanity's existence, social media platforms, and their existence within neoliberal, capitalist structures, have allowed for conspiracy theories to spread to individuals who may never have previously come across fringe views, due to algorithms aimed at keeping people on platforms for as long as possible to generate advertising revenue. The motto of Silicon Valley was "move fast and break things",[2] but it seems like now the pieces must come together, and we must create meaningful change, for the sake of individuals, and democracy more generally.

In Part I I will provide some background to conspiracy theories, looking at why humans are predisposed to believing in them, how they spread, and the impact of social media. This will include a look at the 'definition problem' that surrounds conspiracy theories, before establishing the harmful nature of a number of specific conspiracy theories, and the polluting effect of their ecosystem on social cohesion, public health, and trust in the media, to demonstrate the importance of tackling the issue.

Part II will then envisage a plethora of solutions: addressing the power of government intervention and social media regulation, the effectiveness of social media 'self-regulation, the importance of education, wellbeing and mental health, traditional journalism and the reshaping the advertising industry.

Part III will then conclude with demonstrating the importance of adopting a multidisciplinary approach as the way forward to reducing the damage inflicted on society by harmful conspiracy theories on social media sites.

---

[1] K Anderson, 'How America Lost Its Mind' The Atlantic, 28 December 2017 available at
https://www.theatlantic.com/magazine/archive/2017/09/how-america-lost-its-mind/534231/
[2] S Frenkel & C Kang, *'An Ugly Truth: Inside Facebook's Battle for Domination'* (Harper Collins 2021).

Table of Contents

Part I: The Problem

Those who believe in conspiracy theories would naturally object to my defining of them as such, as to a believer, a conspiracy theory is not conspiratorial, nor a theory, but a truth, a lens through which all else is interpreted.[3] Whilst some scepticism is rational, and demands for transparency from governments and organisations is warranted in a world where conspiracies have been proven to exist (as delineated from conspiracy theories), there is a territory people can end up in which is "characterized by being hyper-skeptical of… information",[4] distrustful of official accounts, believing ideas that are riddled with incoherence but a strong internalised 'logic', and ultimately can lead to anger and violence.

Numerous factors come into play which leave people susceptible to believing conspiracy theories; from feeling powerless and lacking autonomy, to the occurrence of a global event or

---

[3] A LaFrance, 'The Prophecies of Q' The Atlantic (2020) accessed at
https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/ on
17/01/2021
[4] S Lewandowsky & J Cook, 'The Conspiracy Theory Handbook' (2020) available at http://sks.to/conspiracy

threat which is difficult to comprehend.[5] Endorsement of conspiracy theories usually plays into an individual's assumptions or biases, helps them to fit in with their social group, and provides them with a purpose – to uncover and share the 'truth', which only they and their community are privy to. Conspiracy theories comfort individuals, 'empowering' them with an explanation where there had only been uncertainty, in a way that lacks falsifiability due to its self-verifying nature.[6] Once someone adopts a conspiratorial view, their belief operates defensively in a "feedback loop",[7] creating an informational environment that can be impossible to 'de-bunk' from the outside, as there is always an alternative explanation available internally which explains the incoherence.[8] As such, some have compared the internal logic found in conspiracy theories, with that of religious faith[9] due to their foundations of faith in spite of evidence.[10] Hare's 'Parable of the Lunatic'[11] helps to demonstrate that views, or 'bliks' as he referred to them, can be both externally irrational, whilst strongly held personal and meaningful beliefs, irrespective of the nonsensical nature to outsiders.

Conspiracy theories appeal to our nature for wanting an explanation for the bizarre and unexpected, but provide irrational explanations.[12] Some people's brains operate so as to make them more "attracted to" conspiratorial beliefs than the rest of the population – something referred to by Bruder et al as "conspiracy mentality"[13], and by Enders, Uscinski & Seeling as "conspiracy thinking".[14] This denotes that some people are simply more inclined to believing conspiracy theories than other people, and can even extend to people believing in "mutually contradictory" theories.[15] In particular, Swami et al were able to demonstrate that the greatest predictor in believing a made up conspiracy theory was an individual already being shown to have believed in other such theories.[16]

As such, whilst it may seem encouraging to know that not everyone is as prone to 'going down the rabbit hole' as one other,[17] the focus shifts to reducing the wide-spread exposure of

---

[5] ibid, 4.

[6] Douglas, Sutton & Cichocka, 'The Psychology of Conspiracy Theories' in Current Directions in Psychology Science (2017), 26(6), 538.

[7] J Uscinkski, 'Down the Rabbit Hole We Go!' in Conspiracy Theories and The People Who Believe Them (2018), 6.

[8] S Lewandowsky, G E Gignac & K Oberauar, 'The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science' PLoS ONE, 8(10), 3.

[9] A LaFrance (n 3).

[10] G Bezalel, 'Conspiracy Theories and Religion: Reframing Conspiracy Theories as Bliks' in Episteme (2019) 2.

[11] A Flew & R Hare, *'Theology and Falsification: A Symposium'* (1971) in B Mitchell *'The Philosophy of Religion'* Chapter 1 (1977) pp. 13-18.

[12] I Haimowitz, 'No One is Immune: The Spread of Q-anon Through Social Media and the Pandemic' Centre for Strategic and International Studies, 17th December 2020, available at https://www.csis.org/blogs/technology-policy-blog/no-one-immune-spread-q-anon-through-social-media-and-pandemic

[13] Bruder et al, 'Measuring Individual Differences in General Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire' (2013) Front. Psychol 4, 1-15.

[14] A Enders, J Uscinski, M Seeling et al, 'The relationship between social media use and beliefs in conspiracy theories and misinformation' (2021) Political Behaviour.

[15] Bruder et al (n 13).

[16] Swami et al, "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' (2011) British Journal of Psychology 102(3) 443-463.

[17] Enders et al (n 14).

conspiratorial content before it reaches these individuals.[18] Prior to the mass expansion of social media platforms it would have been harder to accidentally come across conspiracy theories, and almost impossible to come across such a large body of supposed 'evidence' without specific intervention or interest on the part of the individual.[19] Facebook recognised that algorithms can "exploit the human brain's attraction to divisiveness" and that "more divisive content" enhances their time online[20] – sites understand their role in polarising society privately, they know how much power they yield in shifting the minds of people, and therefore society, yet publicly they don't want to recognise their role in facilitating such harms. We must acknowledge that conspiracy theory content is prime for social media sites' engagement algorithms due to its "novel and provoking" nature.[21] Platforms must be held accountable, and a multi-pronged solution is required to appropriately reduce both the spread of conspiracy theories online, and their corresponding harmful effects offline.

The Definition Problem

We must be careful to limit the scope of what we mean when we say 'conspiracy theory' as we go forwards. It must be distinguished specifically from misinformation more generally,[22] due to conspiracy theories involving misinformation, but with not all misinformation necessarily conspiratorial in nature. Additionally, conspiracy theories are a broader narrative view of the world, dividing people into "in-groups and out-groups" as opposed to being mere factual inaccuracy. [23] The scope of this paper must be limited to a definition of harmful conspiracy theories, defined in such a way so as to only restrict that subset of online speech which is truly necessary for the preservation of society.[24] As such, it must be acknowledged that some conspiracy theories have historically provided benefits to society; allowing people to question a lack of transparency, showing up inconsistency in official accounts,[25] and can increase productive debate on a topic.[26] Furthermore, beyond the scope of this paper is the recognition that government and authorities can be one of the largest disseminators of conspiratorial content, and, as such, any recommendations shouldn't allow for governments to define conspiracy

---

[18] Enders et al (n 14).
[19] Anderson (n 1).
[20] Horwitz & Seetharaman, 'Facebook Executives shut down efforts to make the site less divisive' The Wall Street Journal, 26th May 2020, https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499?mod=hp_lead_pos5
[21] M Faddoul, G Chaslot & H Farid, 'A Longitudinal Analysis Of Youtube's Promotion Of Conspiracy Videos' Arxiv (Cornell University: Computers And Society) (6 March 2020) Available At Https://Arxiv.Org/Abs/2003.03318
[22] J Miller, K Saunders & C Farhart, 'Conspiracy Endorsement As Motivated Reasoning: The Moderating Roles Of Political Knowledge And Trust' In American Journal Of Political Science 60(4) 4 October 2016 824-844.
[23] F Farinelli, "Conspiracy Theories And Right-Wing Extremism – Insights And Recommendations For P/Cve' (2021) European Commission Radicalisation Awareness Network, available at https://ec.europa.eu/home-affairs/sites/default/files/what-we-do/networks/radica
[24] Sunstein & Vermeule, 'Conspiracy Theories' (2008) University of Chicago Public Law & Legal Theory Working Paper No. 199.
[25] Jolley & Douglas, 'the social consequences of conspiracism: exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint' (2013) British journal of psychology 105(1) 35-56.
[26] S Mari Et Al, 'Conspiracy Theories And Institutional Trust: Examining The Role Of Uncertainty Avoidance And Active Social Media Use' (2021) Political Psychology.

theories in such a way to grant them power to shut down speech on important topics, where transparency should be provided.[27]

As correctly pointed out by Ucinski, it would be inappropriate to give the power to social media companies to define conspiracy theories in such a way that unduly limits freedom of speech.[28] Instead, it seems more appropriate for there to be inter-governmental agreement about a core set of harmful conspiracy theories, which can't be co-opted by individual governments for autocratic purposes[29] – such as stifling genuine and necessary dissent in the face of flagrant abuse of human rights (a criticism which has been levelled against Singapore's 'fake news law').[30]

However, a subset of conspiracy theories go beyond the externally ridiculously, to something provably harmful to society, instigating violence,[31] civil disobedience,[32] and reducing people's sense of autonomy and engagement in politics.[33] For instance, belief in conspiracy theories has been proven to reduce people's intentions to vote, increases people's prejudice towards other groups, increases the likelihood of not taking public health precautions,[34] and, at its most intense, has been seen as a "powerful recruitment tool for extremist ideologies."[35] Subsequently, many have compared an individual becoming increasingly pre-occupied with conspiracy theories, as a form of radicalisation which can lead to domestic terror and self-inflicted societal exclusion.[36] As such, it is fundamentally important to stop conspiracy theories taking root in individuals, due to the difficulty associated with de-radicalising individuals.[37] Thus, Krekó has persuasively argued that content moderation efforts should be focused on theories that are 1) provably harmful 2) have low levels of plausibility, and 3) are popular/viral in nature.[38] I do, however, recognise the risk posed by an ecosystem where conspiratorial content spreads freely, which means all conspiratorial content should be approached cautiously, focusing on education and social policy alongside moderation.[39]

Recommendation 1: An internationally recognised list of harmful conspiracy theories which social media sites are required to act against to limit the spread of – such as;

---

[27] J E Uscinski, 'Conspiracy Theory Vs Free Speech – Should We Regulate Social Media?' Bennett Institute For Public Policy, Cambridge University, 20 March 2019, Available at https://www.bennettinstitute.cam.ac.uk/blog/conspiracy-theory-vs-free-speech-should-we-regulat/
[28] ibid
[29] Bruder et al (n 13).
[30] E Aswad, 'in a world of fake news, what's a social media platform to do?' (2020) Utah Law Review, 4.
[31] Farinelli (n 23).
[32] W Marcellino et al, 'Detecting Conspiracy Theories on Social Media: Improving Machine Learning to Detect and Understand Online Conspiracy Theories. Santa Monica' (2021) RAND Corporation.
[33] Douglas, Sutton & Cichocka (n 6) 539.
[34] J Kramer, 'Why People Latch On To Conspiracy Theories, According to Science' National Geographic, 8 January 2021, available at https://www.nationalgeographic.com/science/article/why-people-latch-on-to-conspiracy-theories-according-to-science
[35] Farinelli (n 23).
[36] H Allam, 'Right Wing Embrace of Conspiracy is 'Mass Radicalization' Experts Warn' NPR, December 15 2020, available at https://www.npr.org/2020/12/15/946381523/right-wing-embrace-of-conspiracy-is-mass-radicalization-experts-warn
[37] A Cháves, 'How QAnon Conspiracy Theories Spread in my Colorado Hometown' The Intercept, 23 September 2020, available at https://theintercept.com/2020/09/23/qanon-conspiracy-theory-colorado/
[38] P Krekó, 'Countering Conspiracy Theories and Misinformation' in Routledge Handbook of Conspiracy Theories (Routledge 2020), 245.
[39] Farinelli (n 23).

1) QAnon – "A cabal theory"[40] identified as a security risk and domestic terror threat in the USA.[41]
2) Anti-Vaccination – In the wake of COVID-19 and the rise of Measles in countries where it was previously eradicated gives renewed importance to these narratives[42]
3) Holocaust Denial – Denial of the amount killed during, or existence of The Holocaust[43]
4) Climate Change Denial – Denial of climate change and its effects[44]

Social Media's Impact

Whilst conspiracy theories have always spread between individuals and groups,[45] it is evident that the "infrastructure and busines models"[46] of platforms in the "modern information marketplace"[47] has allowed for conspiracy theories to spread with less resistance, finding the people who are more inclined towards their beliefs.[48] As such, irrespective of some academics positing that the relationship between social media and conspiracy theories is "conditional on other individual predispositions,"[49] it is unarguable that social media is now one of the largest recruiting spaces for conspiracy theories. Given the ease of access and shareability on social media, it's easy to see why so many people now get the predominant amount of their news in the transient and impassive form of a social media news feed, or through a recommendation algorithm.[50] It must be recognised that research suggests that social media use does correlate (although we cannot necessarily read the word cause into this) with a significantly higher amount of conspiratorial beliefs by individuals.[51] Having a virtual space with access to millions of individuals, and understanding the enhancing factor of the business structures involved is key to recognising the need to shut down common 'pipelines' and trying to get through to existing conspiratorial communities.

People spend a great deal of time on social media sites. Humans are psychologically prone to recalling information they have been exposed to, and platforms have been shown to produce repetition of content which the brain uses as a substitute for validation. Virality and shares

---

[40] S Lewandowsky & S van der Linden, 'Countering Misinformation and Fake News Through Inoculation and Prebunking' (2021) European Review of Social Psychology.
[41] J Winter 'Exclusive: FBI document warns conspiracy theories are a new domestic terrorism threat' Yahoo News 1 August 2019 available at https://news.yahoo.com/fbi-documents-conspiracy-theories-terrorism-160000507.html?guccounter=1
[42] Jolley & Douglas, 'The Effects of Anti-Vaccine Conspiracy Theories on Vaccine Intentions' (2014) PLoS ONE 9(2), 1.
[43] Evans, 'Anti-Semitism Lurks Behind Modern Conspiracy Theories' The Irish Times, 16 February 2021 available at https://www.irishtimes.com/opinion/anti-semitism-lurks-behind-modern-conspiracy-theories-1.4485495
[44] J Lanier, *10 Arguments to Delete Your Social Media Accounts Right Now* (Vintage Publishing, 2018).
[45] Winter (n 41).
[46] Policy@Manchester, 'Infodemic: tackling conspiracy theories on social media' Politics Home, 29 march 2021 https://www.politicshome.com/members/article/infodemic-tackling-conspiracy-theories-on-social-media
[47] Winter (n 41).
[48] B Zadrozny & B Collins, 'How Three Conspiracy Theorists Took Q and sparked QAnon' NBC News, 14 August 2018, available at https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-qanon-n900531
[49] Enders et al (n 14).
[50] Shu, et al 'Fake news detection on social media: a data mining perspective' (2017) ACM SIGKDD Explorations Newsletter Vol 19 Issue 1, 22-36.
[51] Enders et al (n 14).

"conferring…legitimacy"[52] on something that, if they came across in another context, may seem implausible. As such, it would be advisable for all platforms to introduce similar features to TikTok and YouTube, which remind users to take breaks from the screen, and provides guidance regarding the benefits of not spending too much time online. Additionally, more apps should promote 'digital wellbeing' features such as screen time management, to promote people setting healthy boundaries for themselves to feel in control of their time online.[53]

Whilst manipulation may seem like a strong word, it must be considered within the context of the spread of content on social media, and on individuals' abilities to form opinions autonomously. Free will is ultimately impinged by the high levels of influence exerted by online environments which individuals use at intimate moments in life. In such way, Facebook's supposed ability to gauge emotional impact of content is classed by Susser as "manipulative beyond any doubt."[54] Obviously, outside of social media, people can also be manipulated, but the social media algorithm itself could be considered "manipulative practices."[55] We can exert control over people's online eco-system more than their individual friendships and family exposure. As touched upon later, social media sites have the ability to "harness [their] wealth of information to precisely tailor advertisements that exploit their vulnerabilities" [56] (which are often revealed unintentionally, or through digital surveillance). In such a way, we can harness their wealth of knowledge of the human psyche to stop conspiracy theories spreading and harming society rather than undermining people's abilities to think independently.

Recommendation 2: Require social media sites to implement time management and mental health reminders prominently, and at regular intervals during usage of their site.

How Conspiracy Theories Spread

Conspiracy theories often include "kernels of truth…intertwined" in a narrative of power dynamics and group victimisation, fear, and anxiety.[57] As such, conspiracy theories are posited as containing 'stigmatized knowledge'[58] – something which makes believers feel as though they are privy to a level of understanding and knowledge of the world that the 'uninformed' don't have access to. This enhances their sense of belief and reinforces their communities as the

---

[52] P Noor, 'Does Suppressing Online Conspiracy Theorists Work? Experts Weigh In' 30 July 2020, The Guardian, accessed at https://www.theguardian.com/media/2020/jul/30/qanon-does-suppressing-online-conspiracy-theorists-work

[53] TikTok UK, 'NEW! Screen Time Management and Restricted Mode Features on Tiktok' Medium, 4 April 2019, available at https://medium.com/@TikTokUK/new-screen-time-management-and-restricted-mode-features-on-tiktok-86eb30bcf93d

[54] D Susser, B Roessler & H Nissenbaum, 'Online Manipulation: Hidden Influences in a Digital World' (2019) Georgetown Law Technology Review 4(1).

[55] ibid

[56] ibid

[57] Chaves (n 37).

[58] M Barkun, 'Conspiracy theories as stigmatized knowledge' (2015) Diogenes, 62(3-4), 114-120.

'enlightened'. The sense of "foreknowledge"[59] and shared language game[60] allows the community to be hard to reach, and incredibly strong in nature.

Humans are social creatures, and our connections and status can be enhanced by sharing content online, influencing others.[61] As pointed out by Harford,[62] our views on topics such as global warming and politics matter very little in terms of international impact, but matter greatly in terms of cementing our place in social hierarchy,[63] reducing risks of ostracization, and allowing for enhanced communities based on shared beliefs. In line with social identity theory and people's desire for acceptance and belonging,[64] people are inclined towards shifting towards the views that are accepted within their "immediate social network".[65] As such, our news feeds on platforms such as Instagram and Facebook are likely to reflect our existing social groups, likely to back up our existing thoughts, creating echo chamber effects. These can both incline people towards conspiratorial beliefs, whilst also making it harder for individuals to see beyond their existing beliefs. Lanier describes this as the 'Solitary/Pack switch', suggesting personality can change depending on situation; individually we are more cautious and free-thinking, whilst in a pack, due to their hierarchical nature, people often defer to the views of their group, and reject the views of their perceived enemy.[66] Therefore, "democracy [arguably] fails" when, in order to uphold group cohesion, individuals adopt an outwardly irrational ideological position.[67] Similarly, people's desire for belonging means that posting on social media platforms operates by "constant[ly] dosing…social anxiety",[68] ensuring that people keep trying to reach more engagement, enhancing their anxiety, and keeps people stuck in a cycle of hyper-vigilance (making them more prone towards being in a headspace to believe in conspiracy theories).

Furthermore, due to some social media companies' monopolisation of a specific market of media, and a process called network effects (whereby people are more attracted to join platforms which they already have their connections on), Lanier suggests that the only way individuals can preserve themselves, and to signal discontent leading to change, is to leave sites in protest.[69] This sends a message to sites, whose profits are predicated on increasing amounts of people joining the platform, that increasing measures to make the sites more pleasant to use, are necessary in order for their growth to continue.

---

[59] A LaFrance, 'QAnon: Nothing Can Stop What is Coming' The Atlantic, June 2020, available at https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/
[60] L Wittgenstein et al, *Philosophical Investigations* (Wiley-Blackwell 2009).
[61] Bright, 'The Social News Gap: How News Reading and News Sharing Diverge' (2016) Journal of Communication, 66(3), 343-365.
[62] T Harford, *How to Make the World Add Up* (The Bridge Street Press, 2020)
[63] Koski, Xie & Olson, 'Understanding Social Hierarchies: The Neuraland Psychological Foundations of Status Perception' Soc Neurosci (2015) 10(5), 527-550.
[64] Shu (n 50).
[65] Shu (n 50).
[66] Lanier (n 44) 40.
[67] A Guess & A Coppock, 'Does Counter-Attitudinal Information Cause Backlash? Results From Three Large Survey Experiments' British Journal Of Political Science, 50(4), 1497-1515.
[68] Lanier (n 44) 71.
[69] Lanier (n 44) 22.

Recommendation 3: The government should actively promote campaigns which encourage people to take time off from social media, as well as encouraging users to set their own healthy boundaries.

Part II: Solutions

The spread of conspiracy theories on social media needs to be pushed back against proactively in two directions; firstly, to stop more individuals becoming exposed to conspiratorial content, and secondly, (but less pertinent to this discussion) to de-radicalise individuals who have become part of conspiracy theory communities.

Boosting Wellbeing

The growth of conspiracy communities is founded on delegitimising and disorientating people so they resort to indifference,[70] or become drawn to a compelling conspiratorial version of events which comes with a community of people who endorse the ideas and support one another. Henceforth, indifference, or inducing high levels of uncertainty around an issue, is dangerous, and something that conspiracy theories thrive on. As such, a typical route into conspiratorial thinking is, understandably, people looking for self-help content, health information, and information about confusing current events – people wanting an answer, but this search for an answer can often yield conspiratorial and harmful search results,[71] "where misinformation may persist for long periods without contradiction".[72] Feelings of uncertainty and anxiety surrounding an issue can make an answer, even a conspiratorial one, appealing, creating an illusion of understanding and control.[73] Within a social context in which there is a "cultural preference [against] uncertainty"[74] due to struggling with feelings of discomfort and unease, this can induce a process of "collective sensemaking"[75] which strengthens community ties during a crisis. Furthermore, it allows a group to decide on their stance and collectively push-back against any "sense of culpability for their disadvantaged position"[76] thereafter. Given that many individuals drawn to conspiracy theories feel left out from society, it is understandable that they would feel strengthened by defensively coming together, reinforcing a sense of individual and collective worth.[77]

In this way, a cultural desire to attribute blame, or individual success, onto individuals or groups ignores the way the world can operate with luck and chance pulling events into existence, as opposed to necessitating a malevolent force. In this way, as put forward by Malcolm Gladwell in

---

[70] Policy@Manchester (n 46).
[71] Haimowitz (n 12).
[72] L Bode & E Vraga 'In related news, that was wrong: the correction of misinformation through related stories functionality in social media' (2015) Journal of Communication, 65(6).
[73] Swami et al (n 16).
[74] Mari et al (n 26).
[75] K Starbird, E S Spiro & K Koltai, 'misinformation, crisis, and public health – reviewing the literature (June 25 2020) Social Science Research Council, Media Well, available at http://doi.org/10.35650/MD.2063.d.2020
[76] Douglas, Sutton & Cichocka (n 6) 540.
[77] Bode & Vraga (n 72) 1133.

'Outliers', people should try and become comfortable with feeling uncertain, realising that some things are more a product of chance than conspiracy.[78] It's incredibly natural, and understandable why uncertainty and the allure of a social group would pull people towards conspiracy theories – and people aren't necessarily blameworthy for falling into this communities, or patterns of thinking.[79] Their exposure to these ideas when they were in a time of anxiety and heightened uncertainty means they were pre-disposed to wanting an answer, wanting reassurance and a sense of belonging.[80] As such, debunking conspiracy theories once they're firmly rooted becomes incredibly hard – the ideas form part of them, how they think, how they see the world, who they speak to and spend time with – as such it's painful and "involves challenging someone's identity".[81] This is why conspiracy theories go beyond normal misinformation, or an individual being uninformed, as it becomes tightly bound to a sense of self. Due to the fact platforms focus on increasing time online to increase their advertising revenues[82] – and unfortunately human psychology means that pursuing engagement in an ecosystem where the brain dwells and focuses on negative emotions, means they are amplified, causing individuals to spend longer online to try and combat these feelings - something colloquially referred to as 'doom-scrolling'. People recognise that spending excessive amounts of time on social media can be damaging for how we see the world, and for identity formation, but what can be done to help?

Taking into account that exposure alone can induce uncertainty and feelings of unease, it seems that a core part of dismantling the spread of conspiracy theories on social media is to increase wellbeing in society; to encourage people to talk through their unease, to fund mental health services for those experiencing anxiety which makes them prone to conspiratorial thinking, and fund local community groups to give people who may otherwise be isolated a group of people to speak to, rather than keeping their feelings to themselves. Additionally, it would make sense that any duty of care policy towards content moderation, or algorithmic development on social media platforms should include a provision regarding wellbeing – that if, as has widely been reported to have been possible by Facebook,[83] social media sites can gauge our emotional state, this should be utilised to require them to host less political or recognised harmful conspiratorial content on their timeline, rather than promoting it, which would potentially yield higher advertising revenues for them.

Going further, social media sites have been keen to stress to prospective customers (advertisers) their ability to profile advertising towards specific demographics based on the information input on the site, the amount of time they spend online, and they hobbies, or interests. With that in mind, we should utilise knowledge of vulnerable demographics, and those prone to conspiratorial thinking,[84] labelling them as vulnerable, and obligating sites to boost the amount

---

[78] M Gladwell, '*Outliers*' (Little, Brown and Company 2008).
[79] Jolley & Douglas (n 42).
[80] A Cichocka, 'To counter conspiracy theories, boost well-being' Nature 587, 177 (12 November 2020) available at https://www.nature.com/articles/d41586-020-03130-6
[81] Policy@Manchester (n 46).
[82] Lanier (n 44) 19.
[83] Arthur, 'Facebook emotion study breached ethical guidelines, researchers say' The Guardian, 30 June 2014 https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say
[84] S J Min, 'Who believes in conspiracy theories? Network diversity, political discussion, and conservative conspiracy theories on social media' (2021) American Politics Research, 49(5), 415-427.

of light-hearted and moderated connect onto their timelines, rather than allowing posts which fuel negative emotions and conspiracy theories to flood the domain.

If we have an ability to profile individuals who are identifiably vulnerable to conspiracy theories, we should focus on enhancing their wellbeing, and ensuring their online environments do not induce unnecessary and contrived anxiety and isolation, acting as a natural immunisation against conspiratorial thinking.

It is therefore recognised that "[promoting] a sense of common identity, [boosting] feelings of belonging and meaning"[85] and making people feel less alone and their thoughts are heard are fundamental to stopping an environment in which people are vulnerable to conspiracy theories spreading. As such, this is an important aspect of stopping the spread of conspiracy theories on social media that is often left out of discourse, but should be given a prominent place in any set of policy recommendations

Recommendation 4: Legislate for a duty of care from platforms towards users identified by their existing algorithms (if such one is found to exist) as being vulnerable, or in an emotional state, to adapt the content shown away from news and other pages, and towards posts by friends and family, and on lighter topics – a concept called 'off-ramping'.[86]

Recommendation 5: Government funding of mental health and outreach programmes for those who have been taken into conspiratorial thinking, recognising their radicalising nature and potential to move towards external violence and harassment if left unchecked, damaging them and hurting society. Government support for charities working in this sector.

Recommendation 6: Slow speech down – require sites to take on design features which "ask a user if they really mean to post a particular post" (using algorithmic word detection looking out for words connected with conspiracy theories) and "reminding users of terms of services"[87] they agreed to when signing up for an account. This would follow what Instagram has implemented, where a feature detects when posts may contain offensive or harmful content and suggests reflection. Similarly, Twitter has recently created a feature which prompts users to re-consider their language.

Boosting Traditional Journalism and Media

People expect important news to come to them through their social media feeds.[88] Despite the fact that the "traditional linear news cycle" has become more complicated with the introduction

---

[85] Cichocka (n 80).
[86] Haimowitz (n 12).
[87] L Woods, 'The Carnegie Statutory Duty of Care and Fundamental Freedoms' Carnegie UK Trust, December 2019, available at https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/12/10111353/The-Carnegie-Statutory-Duty-of-Care-and-Fundamental-Freedoms.pdf
[88] Bergström & Belfrage, "News in Social Media: Incidental Consumption and the Role of Opinion Leaders." (2018) *Digital Journalism* 6 (5): 583–598, 591.

of social media and the incidental way that many individuals now consume news,[89] traditional media can certainly play a role in stemming the spread of conspiracy theories on social media. As such, but beyond the scope of my expertise, we must look into changing the "funding incentives" which favour certain types of news over more trustworthy and individually relevant, "local journalism…and longform journalism".[90] Looking briefly to this, it can be seen that newspapers have suffered especially from the move towards digital consumption, with their own advertising revenues falling significantly as a result, which has induced "mergers…[and] heavy cuts in staffing".[91] The print media, who are held to stricter editorial standards than social media posters due to their traditional "agenda setting power",[92] is judged by the public to be more reliable in reporting important news to them, but online eco-systems have caused their exposure to such news to decline, and for articles from less trustworthy sources, likely to peddle conspiracy theory interpretations, to thrive. Traditional journalism sources have been pivotal in exposing issues which may not have otherwise come to public attention such as mass surveillance projects undertaken by national security agencies.[93] If present profit motives don't encourage this content at present, then new ways to encourage this must be developed to increase public awareness about real corruption and conspiracies so people don't resort to conspiracy theories for unofficial explanations.

However, it must be recognised that some traditional media has, unfortunately, played heavily into conspiracy theories, and have become more partisan and extreme in their own content (often with views trickling from social media into more traditional news channels),[94] demonstrating that "legacy media" is in no way a monolithic subset of good intentioned news companies attempting to spread important and relevant information to the public. The fact so many news sources in the UK are owned by partisan figures, who therefore promote a specific politics towards their readership in a supposedly 'unbiased' way, is a large problem in itself, which I do not proclaim to be easily solved.

In this way, an investigation into the invigoration of traditional media, or, at least, a promotion of good journalism must be conducted, with profit incentives addressed, and an innovative approach to ensuring good content reaches vulnerable individuals assessed.

It is important to realise that, when exposed to a variety of sources across the spectrum, individuals "rated mainstream sources" significantly "more trustworthy than…hyper-partisan or

---

[89] ibid, 584.

[90] Baron & Crootof (alongside other participants), 'fighting fake news: workshop report: hosted by the information society project (Yale law School) and the floyd abrams institute for freedom of expression' (2018) available at https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf

[91] Cairncross Review: a sustainable future for journalism, Gov.uk, 12th February 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf 6.

[92] S Meraz, 'the fight for 'how to think': traditional media, social networks, and issue interpretation' (2011) Journalism 12(1) 107-127, 120.

[93] E Macaskill & G Dance, 'NSA Files: Decoded: What the revelations made for you' The Guardian (1 November 2013) available at https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded

[94] Lanier (n 44) 48.

fake news sources".[95] This provides hope that, when pushed, people are able to discern good news sources from more uncertain ones, something which could aid the alteration of algorithms displaying content to individuals online. As conspiracy theories thrive from the spread of more low-quality news sources there's a possibility this could be stemmed by sites "up-rank[ing] content from trusted media outlets" rather than frequently targeting individuals with conspiracy content. This approach would fit within platforms' existing business models, utilising independent evaluations of the trustworthiness of news outlets to decide which content to boost or limit. Furthermore, Facebook has shown support for algorithmic developments, countering covid vaccination misinformation by "downranking"[96] misleading content. They've since announced they will take this further, removing posts containing debunked vaccination information to stem the spread of content leading to physical harm.

Recommendation 7: To combat the way that social media users' are de-sensitised to important news due to the incidental way in which it is consumed, sites should be encouraged to include a pop up at the start of each day linking to reputable news sources.

Recommendation 8: Algorithms should prioritise independently judged "credible" news sources, rather than content that principally drives engagement.

Recommendation 9: Governments should reach out to traditional journalism sources with funding and research grants to investigate methods of increasing their reach and impact in the online environment,[97] recognising that they have struggled to diversify their business models in the technology age.[98]

Government Regulation

Until remarkably recently social media sites have largely been able to self-regulate, or form codes of best practice amongst themselves. However, recently political and social developments have caused many countries to consider larger levels of governmental intervention in the online space. In the UK this comes in the form of the Online Safety Bill.

As it stands, the Online Safety Bill would give legal powers to a regulator (Ofcom) to regulate online communications platforms, and those which display "user-generated content publicly or to a large…audience". The duty covers "reasonably foreseeable harm that occurs to people who are users of a service, and reasonably foreseeable harm to people who are not users of a service".[99] This particularly allows for a focus on addressing harmful eco-systems, but the current provisions do not define harm in such a way that would immediately enable the addressing of

---

[95] G Pennycook & D Rand, 'Fighting Misinformation on social media using crowdsourced judgments of news source quality' PNAS February 12 2019 116(7) 2521-2526 accessed at https://doi.org/10.1073/pnas.1806781116

[96] M Isaac, 'Facebook Says it Will Remove Coronavirus Vaccine Misinformation' December 3 2020, New York Times, accessed at https://www.nytimes.com/2020/12/03/technology/facebook-coronavirus-vaccine-misinformation.html

[97] Bright (n 61).

[98] Lanier (n 44) 56.

[99] J Woodhouse, 'Regulating Online Harms: Briefing Paper 8743' House of Commons Library, 12 August 2021.

conspiracy theories. Despite the fact that it highlights harms such as "harms to national security", "democracy", "terrorism" and "disinformation" which, following on from my earlier discussion of the harm inflicted on society by conspiracy theories and their communities leading to social deterioration would appear to be able to be included, it would be preferable to have conspiracy theories explicitly rather than implicitly included so social media platforms (within their defined "category 1" platforms) have no excuse not to recognise their duty to tackle them.

By way of enforcement mechanisms, Ofcom would be given powers to issue civil fines "of up to £18 million, or 10% of a company's global annual turnover", serve notices on companies requiring a response about their breach of duty, and ability to "publish public notices about proven failure of companies to comply". Similarly, in the US, if we were to recognise the physical and psychological violence that can come from conspiracy theories and conspiracy groups, then it would be able to be brought within the scope of Section 230, and re-orientate the framework in favour of people and society rather than social media sites and their shareholders.[100] I will discuss the effectiveness of these measures later.

Recommendation 10: The Online Safety Bill needs to expand the definition of the word harm to explicitly, rather than implicitly, include conspiracy theories due to their recognisable consequences on society.

Duty of Care

It has been persuasively suggested that platforms should be made responsible for the harm they inflict on society,[101] and made to tackle the harm itself, and face fines when shown to have failed in discharging such a duty sufficiently. In this way, a statutory duty of care[102] (as set out comprehensively in case law such as *Donoghue v Stevenson*[103]) can be seen as a form of "harm mitigation,"[104] as well as redress for exposing individuals, and thus society more widely, to the harmful consequences of the spread of conspiracy theories. As such, the problem of curtailing freedom of speech is addressed through ensuring the list of harmful conspiracy theories which require action is established and agreed upon ahead of time, rather than being left to individual platforms to decide. Therefore, we need to consider the violence stemming from conspiratorial groups broadly, looking at the consequences of radicalisation broadly, rather than requiring individual acts of speech to be inciting violence for content to be taken seriously. Providing defined content to limit the reach of would, therefore, allow Facebook to ban groups and remove content on legal grounds with less platform backlash.[105]

---

[100] Centre for Human Technology, 'Policy Principles' available at https://www.humanetech.com/policy-principles
[101] J Ohrvik-Stott & C Miller, 'A Digital Duty of Care: Doteveryone's Perspective' Doteveryone, February 2019, available at https://doteveryone.org.uk/wp-content/uploads/2019/02/Doteveryone-briefing-a-digital-duty-of-care.pdf
[102] Speller, 'Reducing Harm in Social Media Through a Duty of Care' LSE Blogs, 10th May 2018, available at https://blogs.lse.ac.uk/medialse/2018/05/10/reducing-harm-in-social-media-through-a-duty-of-care/
[103] [1932] UKHL 100.
[104] Woods (n 87).
[105] Centre for Human Technology (n 100).

Recommendation 11: Enhance the definition of harm in the UK Online Safety Bill to include provably harmful conspiracy theories (as outlined above) so social media sites have a recognised duty of care to tackle conspiratorial content online, recognising the provable harm they inflict on individuals, communities, and society generally.

In this way, it is important to limit a definition to provably harmed conspiracy theories, rather than risking excessive take-downs, which the public would see as contravening freedom of expression and could stop positive conversations occurring on harmful conspiratorial topics which could lead to people re-thinking their beliefs. Any legal policy must not breach due process principles under the rule of law. Legal duties of care imposed on social media platforms must, therefore, correspond to addressing the ecosystem itself beyond mere content moderation principles – imposing requirements surrounding design features which "hyper-nudge"[106] behaviour for engagement (and profit) incentives.[107] Thus, Woods' duty of care regime, which goes beyond content moderation to structural requirements for monitoring foreseeable harms, requiring mitigation against their propagation on platforms is "risk-based and outcomes-focused", and could re-frame the priorities of social media companies due to the possibility of fines imposed. Until now, "neoliberal technology" and growth driven purely by profit and a lack of external oversight has allowed individuals to become "narrow-minded and easily manipulated".[108] At present, the lack of regulation has allowed for technology companies to consistently put profits and growth before the people using the platforms, with a lack of regard for the global problems they are contributing to. Their priorities need to be shifted more forcibly, as self-regulation has only been able to go so far. Governments must push back against lobbyists to ensure that the good of their people rather than technology companies is being pursued.[109] Neoliberal free market economics means that strong governments must resist pressure from technology corporations to stop positive regulation that would dent company profit margins.[110]

A duty of care must recognise that the overall business models in which these social media sites operate largely drives the spread of harmful conspiracy theories on their platforms, and therefore this eco-system itself needs to be looked at as part of the solution beyond the creation of a narrow content moderation duty.[111]

---

[106] Susser (et al) building on R H Thaler & C R Sunstein, *'Nudge: Improving Decisions About Health, Wealth, and Happiness'* (Penguin 2008).

[107] G Smith, 'A Ten Point Rule of Law Test for a Social Media Duty of Care' CyberLeagle, 16 March 2019, available at https://www.cyberleagle.com/2019/03/a-ten-point-rule-of-law-test-for-social.html

[108] B Bollen, 'Neoliberal Technology is Shaping our World. Let's Turn the Tables on It' Compass Online, 5 January 2017, available at https://www.compassonline.org.uk/neoliberal-technology-is-shaping-our-world-lets-turn-the-tables-on-it/

[109] Institute for the Future, 'Building a Healthy Cognitive Immune System: defending democracy in the disinformation age' (2019) available at https://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_ODNI_Cognitive_Immunity_Map__2019.pdf

[110] D Harvey, 'Neoliberalism as Creative Destruction' (2007) The Annals of the American Academy of Political and Social Science 610, 34.

[111] Ohrvik-Stott & Miller (n 101).

Recommendation 12: A singular regulatory body with one aim, empowered with looking into online conduct would better serve combating the spread of conspiracy theories, and dealing with other online harms, rather than relying on Ofcom, when they have other responsibilities to direct their attentions to.

Recommendation 13: Encourage the UK government to co-ordinate with other governments around the world to create a multi-national regulatory framework to approach conspiracy theories,[112] looking to the work of UNESCO and the EU on the matter, whilst incorporating a duty of care framework. Whilst this may be difficult to reach initially, an international consensus regarding the harms of conspiracy theories will eventually be necessary if we wish to address this problem globally.[113]

Fines and Enforcement

I agree with the UK government regarding their enforcement mechanism backing the Online Safety Bill. A 'pollution' payment rationale towards fines makes sense – for governments to pass on a cost for the harm inflicted on society to those who have benefited – in this case social media platforms.[114] The incentives themselves must point in favour of the interests of people – ensuring that society is not hurt from the spread of conspiracy theories on social media.[115] It goes beyond fining companies, and recognises the harms caused by conspiracy theories online and their overspill into real-life violence.[116] Additionally, the inclusion of publishing notices of insufficient action, or breaches by social media companies, by Ofcom, appears to be incredibly useful for discouraging behaviour – with reputational loss being an "important deterrent to misconduct".[117] For instance, after the Cambridge Analytica scandal,[118] Facebook faced backlash for having received a fine of only £500,000 as many believed this was disproportionately low for the harm suffered. Despite this, the stock price was hit drastically, with the market value having fallen more than $36 billion.[119] One issue is that, whilst reputational damage may be sufficient to cause companies to change their approaches, the monetary damage suffered by the company does not directly go back into improving the harmed society. Therefore, in order for regulation to actually contribute towards the stemming of conspiracy theory content on social media, there must be enforcement mechanisms which are punishable by consistent and transparent fines, which is then invested in education policies (something discussed below in depth). Also, we must

---

[112] I Khan, 'How can states effectively regulate social media platforms' University of Oxford, Faculty of Law, 13 January 2021, available at https://www.law.ox.ac.uk/business-law-blog/blog/2021/01/how-can-states-effectively-regulate-social-media-platforms

[113] William-Perrin, 'Implementing a Duty of Care for Social Media Platforms' Berggruen Institute, 1 March 2021, available at https://www.berggruen.org/ideas/articles/implementing-a-duty-of-care-for-social-media-platforms/

[114] ibid

[115] Khan (n 112).

[116] ibid

[117] J Armour, C Mayer & A Polo 'Regulatory Sanctions and Reputational Damage in Financial Markets' (2017) *Journal of Financial and Quantitative Analysis, 52*(4), 1429-1448.

[118] C Criddle, 'Facebook sued over Cambridge Analytica Data Scandal' BBC News (28 October 2020) available at https://www.bbc.co.uk/news/technology-54722362

[119] S Rodriguez, 'here are the scandals and other incidents that have sent Facebook's share price tanking in 2018' CNBC (20 November 2018) available at https://www.cnbc.com/2018/11/20/facebooks-scandals-in-2018-effect-on-stock.html

recognise that fines cannot exist in a vacuum, and along with payment of a fine, a commitment to adapting the business model to ensure the continuation of a practice is stopped, is required.[120]

Recommendation 14: Ensure that fines are transparent, and big enough to act both as a deterrent, or as retribution after a wrongdoing, with the money redistributed into the community.

Support for Targeted Innovation

Content moderation across social media platforms can have "significant mitigating effects" on the "diffusion of conspiracy theories".[121] However, for this to be facilitated we would need advancements in contextual natural language processing[122] so moderation does not automatically take down helpful discussion, focusing only on specific instances of conspiracy content promotion or recruitment.[123] Due to the volume of content posted daily on social media sites it would be futile, and economically unviable, to require all content be subject to human moderation prior to approval,[124] made ever more futile due to the ever-changing meaning of words within conspiracy communities.[125] At present, therefore, it would be difficult to subject conspiracy content to intensive moderation policies. This is particularly prevalent for content shared through photos or videos, where determining whether content is conspiratorial proves especially hard. As such, investment and research into development, encouraging co-operation between "social science and domain experts",[126] would be useful so platforms can have productive notice and takedown packages governed by realistic requirements. This would also aid the facilitation of the rule of law and due process principles as it would focus attention on only taking down harmful content. This could involve a deeper look into the work of Marcellino et al, who discovered the effectiveness of a "Hybridge approach of stance and deep neural network word embeddings" to "capture genre features", allowing for better classification of content, "reduc[ing] false positive rates", enhancing freedom of speech across valid conversations, and aiding better content moderation.[127]

Recommendation 15: Encourage cross-industry discussion and innovation in the arena of natural language processing focusing on consistency and transparency.[128] Platforms should

---

[120] G Benjamin, 'Gaps in UK regulation of online platforms make it difficult to tackle systemic issues – here are some ways we can fix this' LSE Blogs, 23 September 2020, available at https://blogs.lse.ac.uk/medialse/2020/09/23/gaps-in-uk-regulation-of-online-platforms-make-it-difficult-to-tackle-systemic-issues-here-are-some-ways-we-can-fix-this/

[121] O Papakyriakopoulos et al, 'The Spread Of COVID-19 Conspiracy Theories On Social Media And The Effect Of Content Moderation' Harvard Kennedy School Misinformation Review, 18 August 2020, Available At https://misinforeview.hks.harvard.edu/article/the-spread-of-covid-19-conspiracy-theories-on-social-media-and-the-effect-of-content-moderation/

[122] M Bates, 'Models of Natural Language Understanding' (1995) Proc Natl Acad Sci USA 92(22), 9977.

[123] Marcellino et al (n 32).

[124] S Shahsavari, et al, 'Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news' (2020) Journal of Computational Social Sciences 3, 279-317.

[125] Wittgenstein et al (n 60).

[126] Marcellino et al (n 32).

[127] ibid

[128] C Carlson and H Rousselle, 'Report and Repeat: Investigating Facebook's Hate Speech Removal Process' (2020) available at https://journals.uic.edu/ojs/index.php/fm/article/view/10288/8327

proactively research contextual content moderation which does not rely on users' flagging content (as exposure to conspiracy theories to then flag the content could, itself, prove harmful to individuals who are prone to conspiratorial thinking).[129] Enhancing natural language processing algorithms, coupled with specific conspiracy theories to look for, would stop human moderators having to go through copious amounts of harmful content themselves,[130] and would make moderation requirements practical, and therefore legitimately enforceable. Furthermore, given the unlikely nature of individuals flagging content within conspiracy community groups,[131] a reliance on flagging is inappropriate to stop the spread of conspiracy theories.

Education Policy and Algorithmic Adaptation

Education should play a pivotal role in stemming the spread of conspiracy theories, starting with individuals, towards creating a culture where questioning and uncertainty surrounding global events doesn't lead towards a mass spread of conspiracy thinking within a population. This builds towards societal 'cognitive immunity,'[132] whereby people engage actively with the news (approaching online media and partisan media with a healthy amount of scepticism), so as to participate in society productively. This vision of the human mind as flexible and open to new ideas and explanations is at the core of allowing individuals who believe in damaging conspiracy theories to interrogate their beliefs in light of new evidence and change their minds without being judged for the process they have gone through. As such, I support Vitriol's statement that conspiracy thinking can be discouraged through encouraging the idea that "it's rational to change one's mind in the face of new information".[133] In accordance with Wu, we must focus on freedom of expression and freedom of the press being utilised to 'defend democracy,'[134] ensuring everyone can participate in civic life, and for their own fulfilment, without manipulation.

There are many different approaches to education which may approve effective, and I will explore the effectiveness of a few possible approaches across the next ___ pages.

It is important to also note that, when I use the term education, I am careful not to suggest that traditional academic education is synonymous,[135] for I am concerned with a more specific inoculation, focusing on the specific inclusion of education about conspiracy theory mechanisms and disinformation techniques, so individuals are equipped with the tools necessary to stop themselves being drawn to alternative explanations.

Governmental Policy and Public Policy

---

129 ibid

130 T Simonite, 'Facebook's AI for Hate Speech Improves. How Much Is Unclear' Wired, 5th December 2020, available at https://www.wired.com/story/facebook-ai-hate-speech-improves-unclear/

131 P Ball & A Maxmen, 'The Epic Battle Against Coronavirus Misinformation and Conspiracy Theories' Nature, 27 May 2020, available at https://www.nature.com/articles/d41586-020-01452-z

132 Institute for the Future (n 109).

133 J Kramer (n 34).

134 T Wu, 'Disinformation in the Marketplace of Ideas' (2020) Seton Hall Law Review 51(1) available at https://scholarship.shu.edu/cgi/viewcontent.cgi?article=1758&context=shlr

135 Lanier (n 44) 50.

We should be aiming to curate a culture which encourages people checking their pre-existing biases. As such, the UK should adopt something akin to the European Commission's Information graphics which help individuals to identify the tenants of conspiracy theories in digestible ways; checking the provenance and purpose of sources and attempting to ask people to decide what the tone and balance of an article is.[136] Government bodies should use their educational arms to tap into the education of the young, and to reaching more isolated (and therefore more conspiracy theory vulnerable) individuals.

At present, a Stanford study showed that 82% of 'middle-school[ers]' were "unable to delineate between virality and provenance",[137] illustrating that school policy is a core way of getting information evaluation techniques, and appropriate levels of scepticism, into the general population. Described as 'media literacy', the UK government has recognised the need to embed disinformation recognition from early-on. This should specifically be targeted towards conspiracy theory ideation and techniques beyond dis- and mis- information more broadly, as the techniques can be specific in nature. In the words of Read, a teacher, "we owe it to our students to put conspiracy theories and their debunking on the curriculum."[138] I agree. This should include adding into the curriculum a need to teach students to interactively assess the reliability of sources, and ensure information is verifiable. Government policy should look at endorsing schemes such as 'MediaWise',[139] which have been able to enhance people's abilities to distinguish fact from fiction. We must always remember that, key to stopping the spread of conspiracy theories on social media, is providing individuals (from an early age) with the tools not treat to treat conspiracy theories with legitimacy, and not to share them further. Additionally, evidence points to educators recognising their role in being able to stem the spread of conspiracy theories, utilising non-confrontational discussion in the classroom around pre-approved topics within the conspiracy field would enable teachers to support the facilitation of policy in reducing conspiracy thinking societally. One example to follow could be Finland, who's education system has shown the positive effects of teaching students disinformation tactics, such as how images can be altered,[140] and statistics utilised disingenuously, increasing children's "resistance to fake news".[141] Currently UK strategies about increasing awareness of techniques has proved vague and less-directed, with insufficient attention paid to disinformation techniques across different

[136] C Edmond, 'Conspiracy theories have flourished during the pandemic – here's how to stop them in their tracks,' 02 Sept 2020, World Economic Forum https://www.weforum.org/agenda/2020/09/conspiracy-theories-prevent-spread-covid-19-unesco/

[137] Shellenbarger, 'Most students don't know when news is fake, Stanford study finds' November 21 2016, The Wall Streat Journal https://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576

[138] Read, 'how can we tackle conspiracy theories in schools' tes, 28th march 2021 https://www.tes.com/news/how-can-we-tackle-conspiracy-theories-schools

[139] Dyakon, 'poynter's mediawise training significantly increases people's ability to detect disinformation, new Stanford study finds' December 14, 2020, Poynter, https://www.poynter.org/news-release/2020/poynters-mediawise-training-significantly-increases-peoples-ability-to-detect-disinformation-new-stanford-study-finds/

[140] Henley, 'How finland starts its fight against fake news in primary schools' The guardian 29 January 2020 https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news

[141] Lessenski, 'Just think about it. Findings of the media literacy index 2019' (November 2019) open society institute Sofia, Policy Brief 55 https://osis.bg/wp-content/uploads/2019/11/MediaLiteracyIndex2019_-ENG.pdf

subjects and contexts.[142] At present "only 2% of children in the UK have the critical skills"[143] necessary to decipher fake news, providing an argument for the importance of putting digital literacy on the curriculum so children are less receptive to conspiratorial content they'll come across online.

Recommendation 16: Put digital literacy and teaching about conspiracy theory techniques on the national curriculum.

Recommendation 17: Recognising that older demographics are particularly prone to being taken in by conspiracy theories,[144] more research should be conducted on reaching this demographic by educational policies.

'Pre-Bunking' (inoculation theory)

Pre-bunking[145] provides people with a small dose of conspiracy theory ideation[146] and then immediately provides refutation and an understanding of how the technique worked (such as appealing to emotions, or stressing urgency), so individuals are best-placed to handle exposure to conspiracy theories on social media.[147] Many have recognised the similarity, and utilisation of conspiracy thinking within the sphere of terrorism, explaining the difficulty with attempting to de-radicalise someone.[148] Therefore, education policy must be utilised to stop conspiracy theories spreading and taking root, due to the recognised difficulty in getting individuals out. Given that the brain can mistake familiarity for truth,[149] and the way algorithms are prone to repeating certain content it believes will engage us, it is important that people are, themselves, a wall of defence to stop the spread of conspiracy theories to preserve themselves, and to stop sharing conspiratorial views with those who receive their posts, inducing a form of 'herd immunity'.[150] Getting people to stop sharing content when they deduce its lack of validity, and recognising the deep psychological impact conspiracy theories have on people's attitudes towards society and life generally means that the best thing to do to stop the spread of conspiracy theories is to halt others being exposed.[151]

---

[142] Department for Digital, Culture, Media & Sport, 'Policy Paper: government response to the Cairncross Review: a sustainable future for journalism', Gov.UK, 27 January 2020, Paragraph 29 https://www.gov.uk/government/publications/the-cairncross-review-a-sustainable-future-for-journalism/government-response-to-the-cairncross-review-a-sustainable-future-for-journalism

[143] National Literary Trust, 'commission on fake news and the teaching of critical literacy skills in schools' (2018) https://literacytrust.org.uk/policy-and-campaigns/all-party-parliamentary-group-literacy/fakenews/

[144] Guess, Nagler & Tucker, 'less thank you think: prevalence and predictors of fake news dissemination on facebook' (2019) Science Advances, Vol 5 No 1 (Jan 9th 2019).

[145] W McGuire & D Papageorgis, 'Effectiveness of forewarning in developing resistance to persuasion' Public Opinion Quarterly, 26(1) 1962 24-34 accessed at https://doi.org/10.1086/267068

[146] Lewandowsky & van der Linden (n 40).

[147] J Kramer (n 34).

[148] Radicalisation Awareness Network, 'Harmful conspiracy myths and effective P/CVE countermeasures' 02/10/2020, 2 accessed at https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ran-papers/docs/ran_paper_conspiracy-myths-expert-meeting_en.pdf

[149] Lewandowsky & Van der Linden (n 4).

[150] ibid

[151] Jolley & Douglas (n 42) 7.

Games

Recently games such as Bad News[152] and Go Viral![153] have built on the importance of pre-bunking in disarming the spread of conspiracy theories online. "Social impact game(s)" assist by educating in a way which is accessible and fun, whilst "prompting behaviour change."[154] Their interactive nature enables individuals to use their own agency in changing their mind, something which enhances their ability to leave conspiracy theories behind.[155] What has been observed is that, the process of "active inoculation", getting people to consider the reliability of sources in a way which is engaging for participants, "trains [them] to be more attuned to specific deception strategies"[156], with those who were the most prone to believing false information actually benefiting the most from the exercise. Go Viral!'s focus on educating about how conspiracy theory content can gain huge traction due to its 'shareability', and looking at the causes of conspiracy thinking meant that playing the game even once reduced susceptibility to believe false information for up to three months.[157] Subsequently, utilising games within schools could greatly aid the curation of a culture of healthy scepticism, educating people to question the use of emotionally charged language, rather than sharing posts that are just looking for likes and reactions onwards to others.[158]

Recommendation 18: Requirements to take a digital literacy game when registering an account online, and then at regular intervals thereafter, focusing on specific methods of spreading conspiracy theories and misinformation on that site, so individuals know what to look out for.

Encouraging Internal Debate Online

Studies have shown that getting individuals to "actively generate pro- and counter- arguments themselves" is more effective than them being debunked by others, "as internal arguing is a more involved cognitive process",[159] allowing individuals to autonomously re-assess their beliefs. Given that individuals' views, especially those such as conspiracy theories that frequently operate within close social groups, form part of who they are, it is often easier for individuals' to reconsider their own opinions rather than feel as though they are being induced to change their mind by motivated individuals. We must foster a society where individuals are allowed to grow privately, as an open-minded society with individuals allowed to be receptive to new ideas is important in stopping the spread of conspiracy theories in person and online.[160]

---

152 'Bad News' Game accessed at https://www.getbadnews.com/#intro
153 'Go Viral' Game accessed at https://www.goviralgame.com/en
154 J Roozenbeek & S van der Linden, 'Fake News Game Confers Psychological Resistance Against Online Misinformation' (2019) Palgrave Communications 65(5) accessed at https://www.nature.com/articles/s41599-019-0279-9.pdf
155 Funke & Benkelman, 'want to be a better fact-checker? Play a game' Poynter July 18 2019 https://www.poynter.org/fact-checking/2019/want-to-be-a-better-fact-checker-play-a-game/
156 Roozenbeek & van der Linden (n 154).
157 F Lewsey, 'Cambridge Game 'Pre-Bunks' Coronavirus Conspiracies' accessed at https://www.cam.ac.uk/stories/goviral
158 ibid
159 W McGuire & D Papageorgis 'Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterargument' J Abnorm Soc Psychol 63:326–332 (1961).
160 Jerit & Barabas, 'partisan perceptual bias and the information environment' (2012) The Journal of Politics 74(3) 672-684.

One way in which internal debate can be encouraged is by utilising the Cognitive Reflection Test,[161] created to allow people to assess their intuitive thoughts. Sites should encourage individuals to undertake a test and see where they went wrong, realising that their unthinking intuition (which is the headspace most individuals are in when scrolling on social media) is more flawed than their engaged, questioning one. Encouraging people to realise the importance of maintaining an open mind[162] could help them to stay more engaged online when exposed to conspiracy theory content.[163] People need to be stimulated to be deliberative and reconsider ideas, rather than allowing all new information to merely reinforce pre-existing opinions.[164] Therefore, despite certain individuals being pre-disposed to conspiracy thinking, the discovery itself of this pitfall enables them to remember to over-ride pre-existing conspiracy ideation.

Social media algorithms', particularly Facebook's, can be utilised themselves to correcting misperceptions of those exposed to significant amounts of conspiracy theory content with "corrective information".[165] This enables individuals to be less reluctant to new information as it provides them, organically, with contrary ideas to those they presently hold, allowing them to consider their prior beliefs in a non-condescending way. This backs up McGuire's early research concluding that the process of engaging in internal debate facilitates people reconsidering opinions, routing themselves out of conspiracy thinking beliefs. Thus, the concept of "strategic amplification" is a productive way to utilise sites within their existing business models, moving beyond their content neutral "strategic silence"[166] to active participation in the informational environment to stop the spread of conspiracy theories. What appears imperative is that individuals are intentionally exposed to information countering their conspiracy belief on a site similar to that which initially exposed them to it, which stops them from moving "researching" further,[167] and entering more niche (and extreme) environments. Providing more balance to an individual's feed would naturally counter the current incentivisation of sensationalist content amplified by existing business models, and could annoy certain partisan groups, but this appears necessary for change to occur.[168]

Breaking Up Echo Chambers

An issue exists where people's social media feeds are all optimised towards increasing that individual's amount of time on a given platform – in that every individual is seeing a different feed about the world – we get recommended different news platform's takes on the same event, see different people's views on the same topic. It becomes hard for the democratic process to

[161] Frederick, 'Cognitive Reflection and Decision Making' (2005) Journal of Economic Perspectives (Vol 19 No 4) 25-42.

[162] Bialek & Pennycook, 'The Cognitive Reflection Test is Robust to Multiple Exposures' Behaviour Research Methods 50 (2018) 1953 available at https://link.springer.com/content/pdf/10.3758/s13428-017-0963-x.pdf

[163] Swami, Voracek Stieger, Tran & Furnham, 'Analytic thinking reduces belief in conspiracy theories' (2014) Cognition, 133(3), 572-585.

[164] ibid

[165] Bode & Vraga (n 72).

[166] J Donovan, D Boyd, 'Stop the presses? Moving from Strategic silence to strategic amplification in a networked media ecosystem' (2021) American Behavioural Scientist, 65(2) 333-350.

[167] Ball & Maxman (n 131).

[168] P Noor (n 52).

operate when society doesn't operate on a base level of facts, a shared lexicon of things we can agree on as being true or false, from which we can then decide what our opinions are, and who to vote for, accordingly. One cannot 'walk in another's news feed' in the same way as we can all watch the same news channels and read the same handful of newspapers – the amount of content and variety of sources speaking on a matter online is seemingly endless in comparison to the amount of traditional news outlets creating content.[169] Lanier helpfully compares self-adjusting feeds to a hypothetical history textbook, or encyclopaedia, on a topic which is optimised to only reflect our existing or likely opinions on a matter.[170] This is key with conspiracy theories due to the insular communities in which they operate, as individuals are prone to seeing homogenous views on their timelines – falling into an illusion that their perception is more widespread or respected than it is in reality. Therefore, a key proposal to stop the spread of conspiracy theories on social media is to fundamentally re-shape engagement driven algorithms, requiring regulation to stop an individual's feed becoming myopic. This could mean that an individual who is typically shown content by a particular subset of news sources, is shown an increased amount of diverging sources, if only to provide them with reference points for understanding an outsider perspective. Furthermore, given that homogenous groups are more likely to self-reinforce it is important break existing filter bubbles with network diversity, providing people with enforced outsider perspectives to stop the hyper-optimisation of feeds, which provide proximity to views reflecting your own leading to "extreme polarization and violence."[171]

Thus, nudging has long been understood as a useful psychological mechanism to benefit communities and individuals. For instance, whilst, under current conditions, people are effectively nudged towards conspiracy groups through algorithms designed to keep them online for as long as possible, and utilising their mental state and algorithmic understanding of their existing beliefs,[172] platforms can re-claim the power of nudging. One such recommendation is that of "injecting extremists' online spaces with…ads and banners that present content about which members of the group are likely to disagree", breaking up echo-chamber effects.

Recommendation 19: Governments should encourage platforms to research and take greater action towards the break-up of echo chambers online, providing more diverse content to them to give them the chance to see a broader perspective and reflect on their views.

The notion underpinning the First Amendment in the United States, and the granting of freedom of speech and expression in democratic societies, is premised upon the concept that more speech and content helps to provide balance, leading people to a rounded opinion. However, this currently doesn't take into account algorithms which have produced one-sided feeds, leading to radicalisation of thought and allowing conspiracy theories to thrive. Thus, the marketplace of ideas needs to regain legitimacy, with social media sites actively engaged in the de-radicalisation process themselves.

---

[169] Lanier (n 44) 61.
[170] ibid.
[171] Min (n 84).
[172] N Velásques et al, 'Hidden order across online extremist movements can be disrupted by nudging collective chemistry' (2021) Scientific Reports, 11, 9965.

Harnessing Social Connections

The impact on our thoughts caused by desire for status and belonging[173] can be successfully harnessed, as the powers that shift our views towards conspiracy theories can be utilised to shift them away from them. Bode and Vraga's[174] research is particularly encouraging in this respect demonstrating that information from our connections can change our views through a process called "social correction". This more closely resembles a traditional freedom of speech framework, instigating debate, enhancing democratic processes and autonomy, and encouraging consensus.[175] There is even some hope towards conspiracy thinkers because, despite their scepticism towards outsiders or official accounts, receiving information from those who have previously been embedded in the community, or a friend warmly reaching out to them, can be effective at getting through to them. However, it must be recognised that the insularity of conspiratorial communities can make these individuals hard to reach,[176] meaning that social media sites algorithmically providing people with contrary information appears more effective for reaching isolated individuals. Nonetheless, public authorities should encourage individuals to speak with one another on conspiratorial topics, creating an open space for dialogue.[177] As such, calm and reasoned discussion on Reddit has been shown to productively allow people to discuss ideas, allowing them to shift their opinions on contentious topics.[178] Furthermore, their inclusion of a reward for both the individual who changed their view, and the person who persuaded them, could be used as a template for other social media companies that tend to inadvertently reward controversy and virality with rewards (likes, shares, re-tweets).[179]

Additionally, individuals should be encouraged to counter conspiracy beliefs they come across, acknowledging that engaging committed individuals in public conversation can increase the "complexity of conspiracist responses"[180] becoming increasingly less believable to the general population, reducing the shareability of conspiracy content widely online. As such, connecting with others and attempting to debunk conspiracy ideation may be helpful to stop outsiders believing their ideas, but may not assist with helping believers who are distrustful of others from reconsidering their views.

Recommendation 20: Social campaigns should encourage people to reach out to individuals who they believe are starting to believe in conspiracy theories. Creating a society where people

---

[173] Koski (n 63) 548.
[174] L Bode & E Vraga, 'See Something, Say Something: Correction of Global Health Misinformation on Social Media', (2018) 1131.
[175] Wu (n 134).
[176] Bode & Vraga (n 72) 1132.
[177] Marcellino (n 32) 42.
[178] V Heffernan, 'Our best hope for civil discourse online is on…Reddit' Wired, 16th January 2018, available at https://www.wired.com/story/free-speech-issue-reddit-change-my-view/
[179] Davidow, 'Skinner Marketing: We're the Rats, and Facebook Likes are the Reward' (2013) The Atlantic https://www.theatlantic.com/technology/archive/2013/06/skinner-marketing-were-the-rats-and-facebook-likes-are-the-reward/276613/
[180] S Lewandowsky, GE Gignac, K Oberauer (2013) The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. PLoS ONE 8(10), 8.

empathetically challenge their connections' ideas is core to stopping the spread of conspiracy theories online.[181]

However, as touched upon, an algorithmic approach to diversifying feeds is more likely to be an effective mechanism for stemming the spread of conspiracy theories online than relying on friends and family providing 'correcting information', especially as fact checking has been shown to contribute towards "belief persistence".[182] Policies must be founded upon our understanding of the psychology of conspiracy theories and their internalisation to stop them spreading. When we acknowledge the difficulty of reaching isolated individuals, and conspiracy ideation's proximity to radicalisation, we should not rely too heavily on individual's being reached individually. People who believe in conspiracy theories are likely to see deliberate correction as provoking, and because of theories' lack of falsifiability, such conversation could increase, rather than decrease, their commitment to the theory – something called 'backfire effect'[183] (however, it must be acknowledged that, when looked into by academics recently, the strength of this affect has been doubted)[184]. Conspiracy theories become core to a person's vision of the world, so it can be hard for them to let go of them, especially if they have made personal or financial sacrifices for 'the cause'. The fact individuals largely see the information that appears on their newsfeeds as being a product of chance means they can be more receptive to content which appears to them; something called 'automation bias',[185] allowing them to re-consider their views privately.

Recommendation 21: More research should be conducted into the effectiveness of correcting information and the prevalence of 'backfire effects'.

Incentive Structures

Recognising that "fake news generates more retweets and likes than…reliable posts, spreading 6-20 times faster,"[186] a balancing mechanism should be utilised so that, rather than offering users reward in the form of likes and follows irrespective of the quality of their content, social media platforms should provide a reward system to encourage reliability. Some suggestions[187] have included adding a "trust button" which then up-votes content to more people's timelines, offering "positive examples of user-based assessments" (providing people with perks for accurate posts), or using fact-checkers to recognise reliable posts, and give those who consistently post in a truthful way a "reliable user badge" so people can recognise them as trustworthy figures posting worthwhile content. This recognises the pitfalls of present platform

---

[181] Mari et al (n 26).

[182] E Thorson, 'Belief Echoes: The Persistent Effects of Corrected Misinformation, Political Communication' (2016) 33(3), 460-480.

[183] Nyhan & Reifler, 'when corrections fail: the persistence of political misperceptions' Political Behaviour 32(2), 303-330, 2010.

[184] Guess & Coppock (n 67).

[185] Bode & Vraga (n 174) 1133.

[186] T Sharot, 'To quell misinformation, use carrots – not just sticks' Nature 591, 347 (2021) accessed at https://www.nature.com/articles/d41586-021-00657-0

[187] ibid.

design which have been "weaponised by media manipulators" who seek engagement and fame through posting whatever gets the most engagement.

Recommendation 22: Sites could provide a verification symbol rewarding those who frequently post balanced and correct information so individuals can easily see figures who are more legitimate sources.

Social Media Self-Regulation

Alongside government regulation, there are mechanisms which can be utilised by platforms themselves to combat the spread of conspiracy theory content. Therefore, governments could try encouraging the social media industry to become more creative with their approaches towards profit-making. One particularly interesting suggestion, put forward by Lanier, is for platforms to investigate the diversification of their business models to be less reliant on user engagement. For instance, he suggests, they could experiment with the potential for subscription models,[188] as this would change the dynamic of social media use, with users being the people for whom the company should then work for, rather than just shareholders and data collectors, shifting dynamics but preserving profits.

Case Study of Improvement – YouTube

YouTube has provided good evidence that shifting algorithmic prioritisation of content[189] above others has been able to reduce the viewing of content "containing misinformation…by 50%" in the USA.[190] Given that over 70% of content watched on YouTube comes from 'recommendations', it shows the power of countering a "business model that rewards provocative videos with exposure and advertising dollars",[191] coinciding with personalisation for engagement purposes. Instead, when things are shifted, people may spend as much time on the site, but not be driven down conspiracy theory optimised paths. In particular, it has been shown that YouTube has acted as a radicalisation pipeline from "milder content" towards more extreme material over periods of time.[192] One specific example of this can be shown by The New York Time's podcast 'Rabbit Hole',[193] whereby they investigated the political and social radicalisation of a young man, who's isolation and inquiry into self-help content sent him towards increasingly extreme content. However, it must be noted that, since their findings in 2019, YouTube has de-platformed many of the channels which Mr. Cain had come across during his time on the site (in

---

[188] Lanier (n 44).

[189] The YouTube Team, 'Our Ongoing Work to Tackle Hate' YouTube Official Blog, 5th June 2019, available at https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate

[190] E Ellis, 'The Influencer Scientists Debunking Online Misinformation' Wired, 13th November 2019, available at https://www.wired.com/story/youtube-misinformation-scientists/

[191] K Roose, 'The Making of a YouTube Radical' The New York Times, 8 June 2019, available at https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

[192] Ribeiro et al, 'Auditing radicalization pathways on youtube' (2020) In Conferenceon Fairness, Accountability, and Transparency (FAT* '20) 131-141.

[193] The New York Times, 'Rabbit Hole' Podcast Series, 17 April 2020, accessed at https://www.nytco.com/press/introducing-rabbit-hole-a-new-narrative-audio-series-from-the-new-york-times/

total taking down over 25,000 channels due to "supremacist content.")[194] Additionally, they have taken steps to 'nudge' watchers of conspiracy content towards "more authoritative news sources,"[195] something I would recommend for them to continue and expand upon. Going further, their prohibition of advertisements alongside hate content[196] could be extended towards harmful conspiracy theories[197] due to the proven hate pipeline and violence attached to such content. Equally, YouTube have shown ability to implement an "education, documentary, scientific, and artistic content" exception, which could be a template for other platforms, demonstrating an ability to delineate between positive discussion about conspiracy theories, and their promotion. Importantly, however, most of these positive changes appear to come after financial repercussions - advertisers pulling out due to being uncomfortable with content on the site,[198] negative press,[199] and previous workers coming out against the companies' previous practices.[200] As such, despite a "massive reduction in conspiratorial recommendations", there was found to still be a relatively high chance of a conspiracy theory video being recommended immediately after another one, reenforcing filter bubble effects.[201] This suggests that self-regulation has been able to angle the site in the right direction, but hasn't been able to go far enough in removing harmful content.

Recommendation 23: Encourage social media sites to operate forms of self-regulation, recognising that some technology changes at such pace that governments will always be a step behind any developments, for instance by reducing the reach of content when debunked or when the site determines the content to be sufficiently harmful.[202]

Deplatforming

The possibility of utilising deplatforming methods to decrease the spread of conspiracy theories on social media is controversial, but has recently been looked at more seriously by governments and platforms. Research has shown that, during COVID-19, only 12 people were responsible for 73% of the misinformation found regarding vaccines online,[203] so deplatforming small numbers

---

[194] J Alexander, 'YouTube Bans Stefan Molyneux, David Duke, Richard Spencer, and More For Hate Speech' The Verge, 29 June 2020, available at https://www.theverge.com/2020/6/29/21307303/youtube-bans-molyneux-duke-richard-spencer-conduct-hate-speech

[195] Roose (n 191).

[196] YouTube Team (n 133).

[197] The YouTube Team, 'Managing Harmful Conspiracy Theories on YouTube' YouTube Official Blog, 15 October 2020, available at https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/

[198] Roose & Conger, 'YouTube to Remove Thousands of Videos Pushing Extreme Views' The New York Times, 5th June 2019, available at https://www.nytimes.com/2019/06/05/business/youtube-remove-extremist-videos.html

[199] Ibid.

[200] M Bergen, 'YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant" Bloomberg, 2nd April 2019, available at https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant

[201] Faddoul et al (n 21).

[202] Facebook, 'Taking action against people who repeatedly share misinformation' Facebook Newsroom, 26 May 2021, available at https://about.fb.com/news/2021/05/taking-action-against-people-who-repeatedly-share-misinformation/

[203] E Salam, 'Majority of Covid misinformation came from 12 people, report finds' The Guardian, 17 July 2021, available at https://www.theguardian.com/world/2021/jul/17/covid-misinformation-conspiracy-theories-ccdh-report

of people or communities could have a huge positive effect on stopping the spread of conspiracy theories online, and is something I therefore advocate for.

Deplatforming has been effective at reducing the reach of harmful views online, and could certainly be utilised as part of a wider campaign of content moderation to ensure that social media platforms become less radical places. Deplatforming diminishes the reach and amplification of harmful and illegal content online, with extreme commentators and groups being pushed to the periphery of online communication.[204] It decreases influence,[205] decreases perceived outsider legitimacy, and therefore decreases the reach and the ability for others to be drawn accidentally (or as a product of engagement boosting algorithms) into conspiracy theory rhetoric.[206] One important impact of deplatforming is that it also renders it almost impossible for large communities to discuss co-ordination and recruitment unless they re-centralise elsewhere.

One example of effective de-platforming utilised by social media sites was especially drawn into media focus following the January 6th Capitol Riots, after which 70,000 accounts were suspended from mainstream platforms due to their connection to harmful right-wing conspiracy theories, which were connected with insurrectionist intentions and political violence.[207] Additionally, in the UK, Facebook's removal of the Britain First page significantly "disrupted the group's online activity", pushed them to less used platforms with less users (a positive effect of network effects as the majority of individuals choose to remain on larger sites where their friends and colleagues are, which are more moderated), and reduced their followers from over 2 million to under 19,000 individuals.[208] Twitter also changed their approach towards favouring deplatforming over time, moving away from wholesale 'free speech',[209] towards the 2021 removal of accounts following the events in January. Ultimately, I fully endorse the sentiment of Romano who has suggested that our recent viewing of technology and social media companies utilising de-platforming and content moderation is an "attempt to prioritise the public good over extremists' demand for a megaphone"[210] which can only be a good thing.

Recommendation 24: Social media platforms should continue to de-platform groups posting recognisably harmful conspiracy theories, and ensure that individuals must recognise, when registering, that the site does not tolerate such content. Governments should aim to collaborate with platforms in understanding the best ways to support their ongoing practices.

Recommendation 25: Look into the possibility of bringing statutory deplatforming provisions into law with a recognised appeals process.

---

[204] Noor (n 52).

[205] A Romano, 'Kicking people off social media isn't about free speech' Vox, January 21 2021, available at https://www.vox.com/culture/22230847/deplatforming-free-speech-controversy-trump

[206] L Nouri, N Lorenzo-Dus, A Watkin, 'Following the Whack-a-Mole: Britain First's Visual Strategy from Facebook to Gab' Global Research Network on Terrorism and Technology Paper No. 4

[207] Romano (n 205)/

[208] J Mulhall, 'Deplatforming Works: Let's Get on With It' Hope Not Hate, 4th October 2019, available at https://www.hopenothate.org.uk/2019/10/04/deplatforming-works-lets-get-on-with-it/

[209] Halliday, 'Twitter's Tony Wang: We are the free speech wing of the free speech party' The Guardian, 22 March 2012, available at https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech

[210] Romano (n 205).

Recommendation 26: Social media sites should monitor those who are increasingly spreading disinformation, hate speech, and conspiracy theory content as "key spreaders."[211] In order to contribute to the rule of law and digital due process, sites should come up with a threshold over which individual users or groups cannot cross before they will be de-platformed, and this should be published in the terms of service so people cannot complain of a lack of knowledge on policy.

Push-Back

Some people, however, disagree with deplatforming, with there being two main lines of argument

1) That it's ineffective
2) That it constitutes an infringement of freedom of speech rights as guaranteed under Article 10 in the EU (incorporated into UK law as Article 10 of the Human Rights Act) and the First Amendment in the US.

Based on the information presented in the paragraph prior, I believe that the argument that it lacks effectiveness is plainly incorrect. Addressing free speech concerns, it is important to remember the functions of speech[212] in optimising society, allowing for conversation and political and civic communication – not the kind of speech which occurs in filter bubbles or includes hateful content or that which would undermine society's functioning.[213] Thus, it appears to me that counter-speech is currently insufficiently tackling harmful content and conspiracy online, especially because minorities often have less social capital,[214] meaning the marketplace of ideas often upholds existing ideas and structures, as those with most impactful voices are usually those who have already benefitted from the maintenance of existing power structures

Furthermore, it must be acknowledged that, irrespective of whether people like the fact that social media platforms have such a say in determining the speech allowed in the online 'public squares', the reality is that they do,[215] and that they are private companies with first amendment rights of their own.[216] As such, unless governments expressly intervene, platforms can utilise deplatforming practices to allow their site to align to their values, and stop them from receiving bad press.

In this way, some have pointed to the fact that de-platforming individuals from social media makes them more prone towards doubling-down harder on their conspiracy views, away from mainstream platforms with more possibility of encounters with outsiders, migrating to less regulated platforms such as Parler, Telegram, and Whatsapp. I do recognise the importance of engaging empathetically with radicalised individuals so they are not deemed 'unsavable', and left

---

[211] Shu, et al (n 50).
[212] N Cofnas, 'Deplatforming Wont Work' Quillette, 8th July 2019, available at https://quillette.com/2019/07/08/deplatforming-wont-work/
[213] Mulhall (n 208).
[214] Ibid.
[215] Romano (n 205).
[216] J York, 'Users, not tech executives, should decide what constitutes free speech online' MIT Technology Review, 9th January 2021, available at https://www.technologyreview.com/2021/01/09/1015977/who-decides-free-speech-online/

out of society,[217] only communicating in less moderated platforms.[218] Similarly, it has been argued that platforms announcing censorship of specific views could give it a veil of secrecy and legitimacy (or argument that the platform themselves are 'in on it')[219] that would be picked up on by conspiracy communities as evidence for their validity.[220] However, in answer to both these criticisms, ensuring these ideas are not readily available on mainstream platforms is still useful in itself, with the migration of conspiratorial individuals reducing the reach of harmful content, therefore helping to stop the spread of conspiracy theories online. Furthermore, addressing Cofnas' concern about allowing social media sites' to become arbiters of truth, I only suggest utilising defined categories of harmful conspiracy theories agreed upon by think-tanks, governments, and social media sites collaboratively, not taking down all content. Therefore, more research into how to stop these negative effects, which I duly acknowledge, would be useful going forward, but is outside of the scope of this academic piece.

One other recommendation to implement in tandem with deplatforming is marginalisation strategy – which attempts to recognise the huge task of attempting to take down all illegal, harmful, and objectionable content online, and instead focuses on isolating them online, and making it easier for law enforcement agencies to "detect, monitor, and investigate"[221] concerning activity. What's key to marginalisation strategy's possible applicability to reduce the spread of conspiracy theories is that it aims to make groups decide between platforms with less reach (and less recruitment potential), but higher security and freedom of speech. It also benefits from allowing the decision to be taken by groups themselves (with social media sites making it harder to operate on openly so they seek out an alternative), as opposed to relying wholly on take-downs themselves (which are reactive and costly to keep doing).[222] As such, this has been shown to be a useful method by counter-terrorism organisations, who focus on "deprecating the influence of violent extremists by progressively undermining, drowning out, and side-lining radical perspectives".[223] A definite pro to this approach is that it has been shown to work, and allows for detection agencies to find the most extreme actors, and focuses on limiting the reach of conspiracy messaging, and reducing the number of vulnerable individuals on platforms who come across conspiracy theory content.

Recommendation 27: For law enforcement agencies, counter-terrorism organisations, and social media sites to come together to form a co-ordinated approach which helps to marginalise conspiratorial communities and actors from large platforms.

---

[217] Cofnas (n 212).
[218] Romano (n 205).
[219] Noor (n 52).
[220] Cofnas (n 212).
[221] B Clifford & H Powell, 'De-platforming and the Online Extremist's Dilemma' Lawfare Blog, 6th June 2019, available at https://www.lawfareblog.com/de-platforming-and-online-extremists-dilemma
[222] A Alexander & W Braniff, 'Marginalizing Violent Extremism Online' Lawfare Blog, 21st January 2018, available at https://www.lawfareblog.com/marginalizing-violent-extremism-online
[223] ibid

Recommendation 28: For the UK, EU, and US to work together to better understand smaller platforms which de-platformed conspiracy theorists have migrated to and figure out ways to reach these individuals in a productive way.[224]

Fact Checking

Fact checking on social media sites can prove unhelpful in dissuading conspiracy theorists from their view, due to their ability to explain inconsistencies and external knowledge, and because of the scepticism towards mainstream media organisations and their coverage of the news. It is also imperative to remember that exposure to conspiracy theories alone can put doubt and uncertainty into an individual's head, so attempts to undermine a conspiracy theory can "inadvertently aid [the creation of] familiarity with incorrect beliefs".[225] Therefore, journalistic sources and platforms fact checking can increase the reach of conspiracy theories despite good intentions,[226] and can lend a more renowned journalistic name to a theory in the mind of individuals.[227]

Recommendation 29: Not to advocate for fact-checking articles or posts to be boosted, but instead for diverse information and counter ideas to be displayed, so as to not amplify a harmful conspiracy theory.

Advertising

Recently there has been more emphasis on utilising advertising as a means to nudge platforms to take action on important social issues that matter to their clients. For instance, Lanier has suggested that it was only when advertisers complained to Google regarding their content being placed alongside terrorist recruitment content that YouTube intervened[228] – showing the importance of utilising commercial pressure to stem conspiracy theories online. Going further, the Stop Funding Hate Campaign of 2020 identified the potential power held by advertisers in sending a message via their profit margins and public image. Given that advertising generates 98% of Facebook's revenue,[229] we must not underestimate the impact that co-ordinated action by advertisers (sites' "real customers"[230]) can have on platforms. Equally, AT&T had pulled their advertising content from YouTube for 2 years between 2017 and 2019 due to content "appearing near offensive content too often".[231] This shows how companies, concerned with preserving their brand image, can proactively demand change in an ecosystem increasingly attuned to targeted advertising, and aware of the 'ethics of influence'.[232] People have gradually

---

[224] Mulhall (n 208).

[225] Kramer (n 34).

[226] Thorson, (n 182).

[227] Enders et al (n 14).

[228] Lanier (n 44) 52.

[229] The Economist, 'facebook eyes a future beyond social media' The Economist, 29th July 2021, available at https://www.economist.com/business/2021/07/29/facebook-eyes-a-future-beyond-social-media

[230] Lanier (n 44) 5.

[231] Maheshwari, 'AT&T to advertise on youtube again after a nearly 2-year holdout' The New York Times, 18th January 2019, available at https://www.nytimes.com/2019/01/18/business/media/att-youtube-advertising.html

[232] C Sunstein, '*The Ethics Of Influence: Governments In The Age Of Behavioural Science*' (Cambridge University Press 2016).

become more aware, especially following the Cambridge Analytica scandal, that online advertising has the ability to emotionally target us, and therefore pull us in certain directions.[233]

As recently as January of 2021, following in the wake of the attempted insurrection at the US Capitol, Facebook were accused of "microtargeting extremists towards weapon accessories" [234] when they ran advertisements regarding military equipment alongside posts which shared "election misinformation…and news about the attempted coup".[235] Thus, we must recognise that platforms' algorithms are optimised for increasing advertising views and revenue, [236] which has been key to driving conspiracy theory content online to date as engagement metrics often "amplify sensational content" [237] to keep individuals online. Thus, to remove harmful conspiracy content, we need to look at advertising regulation's potential and the possibility of companies taking their own steps to address any failings if regulation fails to protect their interests. Therefore, the recommendations outlined below aim to mitigate the effects of individual targeted advertisements, as well as the eco-system induced on social media sites to increase advertising views.

Recommendation 30: Sites should be encouraged to engage in conversation with advertisers' about their desires for content removal and their concerns about being associated with platforms which promote conspiracy content, and brands who feel strongly on such issues should pull their content from the site if no change occurs.

Recommendation 31: To find and remove both "public and private groups focused on" harmful content, including "violent conspiracies".[238]

Recommendation 32: Governments should investigate regulating the advertising industry as a way to reach social media companies', directing them towards reducing harmful content online.

Recommendation 33: Platforms should be required to more clearly differentiate advertised posts so user can easily delineate between content from friends, family, pages they like, and suggested content they're likely to engage with, and posts specifically targeted towards them (known as "native advertising"[239]).

Conclusion: The Way Forward

---

[233] Susser et al (n 54).

[234] Mac & Silverman, 'Facebook has been showing military gear ads next to insurrection posts' Buzzfeed News, January 13 2021, available at https://www.buzzfeednews.com/article/ryanmac/facebook-profits-military-gear-ads-capitol-riot

[235] ibid

[236] Policy@Manchester (n 46).

[237] Faddoul et al (n 21).

[238] Stop Hate For Profit, 'One year after stop hate for profit: progress update from the coalition' Stop Hate for Profit (30 July 2020) https://www.stophateforprofit.org/

[239] Susser et al (n 54).

I have put forward many individual recommendations, but if I were to provide an overarching method for combating the spread of conspiracy theories it would be for the government to encourage a divergence from existing models where profit has been able to be derived from enhancing engagement at all costs, something which has been a constant driver of the spread of provocative and harmful misinformation. We need individuals to be at the front and centre of social media; the people who platforms must serve, rather than be exploited by. The market of online manipulation must be disrupted, and individuals must be encouraged to form communities, for individuals to look out for vulnerable family members and friends who display the hallmarks of conspiracy thinking. The problem is the eco-system rather than tackling individual groups and online communities.

I have no desire to blame individuals, who are often vulnerable and experiencing anxiety and stress, but to focus on solutions, to be hopeful that individuals and communities can be understood and unravelled, that communities can come together once again. We need a holistic societal approach encompassing education, regulation, and a culture of wellbeing and openness in order for that to be achieved.

Recommendation 34: Be hopeful – if we take action the spread of conspiracy theories online can be tackled effectively.

Bibliography

Cases
1. Donoghue v Stevenson [1932] UKHL 100

Books

1. A Flew & R Hare, '*Theology and Falsification: A Symposium*' (1971) in B Mitchell *'The Philosophy of Religion'* Chapter 1 (1977) pp. 13-18
2. A Wiener, *Uncanny Valley: A Memoir,* (MCD 2020)
3. C Sunstein, '*The Ethics Of Influence: Governments In The Age Of Behavioural Science*' (Cambridge University Press 2016)
4. J Lanier, *10 Arguments to Delete Your Social Media Accounts Right Now* (Vintage Publishing, 2018)
5. L Wittgenstein et al, *Philosophical Investigations* (Wiley-Blackwell 2009)
6. M Gladwell, '*Outliers*' (Little, Brown and Company 2008)
7. R H Thaler & C R Sunstein, '*Nudge: Improving Decisions About Health, Wealth, and Happiness*' (Penguin 2008)
8. S Frenkel & C Kang, *'An Ugly Truth: Inside Facebook's Battle for Domination'* (Harper Collins 2021)
9. S Lewandowsky & J Cook, 'The Conspiracy Theory Handbook' (2020) available at http://sks.to/conspiracy
10. T Harford, *How to Make the World Add Up* (The Bridge Street Press, 2020)

Journal Articles

1. A Enders, J Uscinski, M Seeling et al, 'The relationship between social media use and beliefs in conspiracy theories and misinformation' (2021) Political Behaviour
2. A Guess & A Coppock, 'Does Counter-Attitudinal Information Cause Backlash? Results From Three Large Survey Experiments' British Journal Of Political Science, 50(4), 1497-1515

3. Baron & Crootof (alongside other participants), 'fighting fake news: workshop report: hosted by the information society project (Yale law School) and the floyd abrams institute for freedom of expression' (2018) available at https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf

4. Bergström & Belfrage, "News in Social Media: Incidental Consumption and the Role of Opinion Leaders." (2018) *Digital Journalism* 6 (5): 583–598, 591

5. Bialek & Pennycook, 'The Cognitive Reflection Test is Robust to Multiple Exposures' Behaviour Research Methods 50 (2018) 1953 available at https://link.springer.com/content/pdf/10.3758/s13428-017-0963-x.pdf

6. Bright, 'The Social News Gap: How News Reading and News Sharing Diverge' (2016) Journal of Communication, 66(3), 343-365

7. Bruder et al, 'Measuring Individual Differences in General Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire' (2013) Front. Psychol 4, 1-15

8. C Carlson and H Rousselle, 'Report and Repeat: Investigating Facebook's Hate Speech Removal Process' (2020) available at https://journals.uic.edu/ojs/index.php/fm/article/view/10288/8327

9. Cairncross Review: a sustainable future for journalism, Gov.uk, 12th February 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf

10. D Harvey, 'Neoliberalism as Creative Destruction' (2007) The Annals of the American Academy of Political and Social Science 610

11. D Susser, B Roessler & H Nissenbaum, 'Online Manipulation: Hidden Influences in a Digital World' (2019) Georgetown Law Technology Review 4(1)

12. Department for Digital, Culture, Media & Sport, 'Policy Paper: government response to the Cairncross Review: a sustainable future for journalism', Gov.UK, 27 January 2020, Paragraph 29 https://www.gov.uk/government/publications/the-cairncross-review-a-sustainable-future-for-journalism/government-response-to-the-cairncross-review-a-sustainable-future-for-journalism

13. Douglas, Sutton & Cichocka, 'The Psychology of Conspiracy Theories' in Current Directions in Psychology Science (2017), 26(6)

14. E Aswad, 'in a world of fake news, what's a social media platform to do?' (2020) Utah Law Review

15. E M Iacobucci, 'On the Interaction between Legal and Reputational Sanctions' (2014) The Journal of Legal Studies, 43(1), 189-207

16. E Thorson, 'Belief Echoes: The Persistent Effects of Corrected Misinformation, Political Communication' (2016) 33(3), 460-480

17. F Farinelli, "Conspiracy Theories And Right-Wing Extremism – Insights And Recommendations For P/Cve' (2021) European Commission Radicalisation Awareness Network, available at https://ec.europa.eu/home-affairs/sites/default/files/what-we-do/networks/radica

18. Frederick, 'Cognitive Reflection and Decision Making' (2005) Journal of Economic Perspectives (Vol 19 No 4) 25-42

19. G Bezalel, 'Conspiracy Theories and Religion: Reframing Conspiracy Theories as Bliks' in Episteme (2019)

20. G Pennycook & D Rand, 'Fighting Misinformation on social media using crowdsourced judgments of news source quality' PNAS February 12 2019 116(7) 2521-2526 accessed at https://doi.org/10.1073/pnas.1806781116

21. Guess, Nagler & Tucker, 'less thank you think: prevalence and predictors of fake news dissemination on facebook' (2019) Science Advances, Vol 5 No 1 (Jan 9th 2019)

22. I Khan, 'How can states effectively regulate social media platforms' University of Oxford, Faculty of Law, 13 January 2021, available at https://www.law.ox.ac.uk/business-law-blog/blog/2021/01/how-can-states-effectively-regulate-social-media-platforms

23. J Armour, C Mayer & A Polo 'Regulatory Sanctions and Reputational Damage in Financial Markets' (2017) *Journal of Financial and Quantitative Analysis, 52*(4), 1429-1448.

24. J Donovan, D Boyd, 'Stop the presses? Moving from Strategic silence to strategic amplification in a networked media ecosystem' (2021) American Behavioural Scientist, 65(2) 333-350

25. J Miller, K Saunders & C Farhart, 'Conspiracy Endorsement As Motivated Reasoning: The Moderating Roles Of Political Knowledge And Trust' In American Journal Of Political Science 60(4) 4 October 2016 824-844

26. J Roozenbeek & S van der Linden, 'Fake News Game Confers Psychological Resistance Against Online Misinformation' (2019) Palgrave Communications 65(5) accessed at https://www.nature.com/articles/s41599-019-0279-9.pdf

27. J Uscinkski, 'Down the Rabbit Hole We Go!' in Conspiracy Theories and The People Who Believe Them (2018)

28. J Woodhouse, 'Regulating Online Harms: Briefing Paper 8743' House of Commons Library, 12 August 2021

29. J York, 'Users, not tech executives, should decide what constitutes free speech online' MIT Technology Review, 9th January 2021, available at https://www.technologyreview.com/2021/01/09/1015977/who-decides-free-speech-online/

30. Jerit & Barabas, 'partisan perceptual bias and the information environment' (2012) The Journal of Politics 74(3) 672-684

31. Jolley & Douglas, 'The Effects of Anti-Vaccine Conspiracy Theories on Vaccine Intentions' (2014) PLoS ONE 9(2)

32. Jolley & Douglas, 'the social consequences of conspiracism: exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint' (2013) British journal of psychology 105(1) 35-56

33. Koski, Xie & Olson, 'Understanding Social Hierarchies: The Neuraland Psychological Foundations of Status Perception' Soc Neurosci (2015) 10(5), 527-550

34. L Bode & E Vraga 'In related news, that was wrong: the correction of misinformation through related stories functionality in social media' (2015) Journal of Communication, 65(6)

35. L Bode & E Vraga, 'See Something, Say Something: Correction of Global Health Misinformation on Social Media', (2018)

36. L Nouri, N Lorenzo-Dus, A Watkin, 'Following the Whack-a-Mole: Britain First's Visual Strategy from Facebook to Gab' Global Research Network on Terrorism and Technology Paper No. 4

37. L Woods, 'The Carnegie Statutory Duty of Care and Fundamental Freedoms' Carnegie UK Trust, December 2019, available at https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/12/10111353/The-Carnegie-Statutory-Duty-of-Care-and-Fundamental-Freedoms.pdf

38. M Barkun, 'Conspiracy theories as stigmatized knowledge' (2015) Diogenes, 62(3-4), 114-120

39. M Bates, 'Models of Natural Language Understanding' (1995) Proc Natl Acad Sci USA 92(22)

40. M Faddoul, G Chaslot & H Farid, 'A Longitudinal Analysis Of Youtube's Promotion Of Conspiracy Videos' Arxiv (Cornell University: Computers And Society) (6 March 2020) Available At Https://Arxiv.Org/Abs/2003.03318

41. N Jones et al, 'Representation of Future Generations in United Kingdom Policy-Making' (2018) Futures 102

42. N Velásques et al, 'Hidden order across online extremist movements can be disrupted by nudging collective chemistry' (2021) Scientific Reports, 11

43. Nyhan & Reifler, 'when corrections fail: the persistence of political misperceptions' Political Behaviour 32(2), 303-330,

44. P Krekó, 'Countering Conspiracy Theories and Misinformation' in Routledge Handbook of Conspiracy Theories (Routledge 2020)

45. P McGorry, T Bates & M Birchwood, 'designing youth mental health services for the 21st century: examples from Australia, Ireland, and the UK' (2013) The British Journal of Psychiatry, 202(54), 30-35

46. Ribeiro et al, 'Auditing radicalization pathways on youtube' (2020) In Conferenceon Fairness, Accountability, and Transparency (FAT* '20) 131-141

47. S J Min, 'Who believes in conspiracy theories? Network diversity, political discussion, and conservative conspiracy theories on social media' (2021) American Politics Research, 49(5), 415-427

48. S Lewandowsky & S van der Linden, 'Countering Misinformation and Fake News Through Inoculation and Prebunking' (2021) European Review of Social Psychology

49. S Lewandowsky, G E Gignac & K Oberauar, 'The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science' PLoS ONE, 8(10)
50. S Lewandowsky, GE Gignac, K Oberauer (2013) The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. PLoS ONE 8(10)
51. S Mari Et Al, 'Conspiracy Theories And Institutional Trust: Examining The Role Of Uncertainty Avoidance And Active Social Media Use' (2021) Political Psychology
52. S Meraz, 'the fight for 'how to think': traditional media, social networks, and issue interpretation' (2011) Journalism 12(1) 107-127
53. S Shahsavari, et al, 'Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news' (2020) Journal of Computational Social Sciences 3, 279-317
54. Shu, et al 'Fake news detection on social media: a data mining perspective' (2017) ACM SIGKDD Explorations Newsletter Vol 19 Issue 1, 22-36
55. Sunstein & Vermeule, 'Conspiracy Theories' (2008) University of Chicago Public Law & Legal Theory Working Paper No. 199
56. Swami et al, "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' (2011) British Journal of Psychology 102(3) 443-463
57. Swami, Voracek Stieger, Tran & Furnham, 'Analytic thinking reduces belief in conspiracy theories' (2014) Cognition, 133(3), 572-585
58. T Wu, 'Disinformation in the Marketplace of Ideas' (2020) Seton Hall Law Review 51(1) available at https://scholarship.shu.edu/cgi/viewcontent.cgi?article=1758&context=shlr
59. W Marcellino et al, 'Detecting Conspiracy Theories on Social Media: Improving Machine Learning to Detect and Understand Online Conspiracy Theories. Santa Monica' (2021) RAND Corporation
60. W McGuire & D Papageorgis 'Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterargument' J Abnorm Soc Psychol 63:326–332 (1961)
61. W McGuire & D Papageorgis, 'Effectiveness of forewarning in developing resistance to persuasion' Public Opinion Quarterly, 26(1) 1962 24-34 accessed at https://doi.org/10.1086/267068
62. William-Perrin, 'Implementing a Duty of Care for Social Media Platforms' Berggruen Institute, 1 March 2021, available at https://www.berggruen.org/ideas/articles/implementing-a-duty-of-care-for-social-media-platforms/

Web-Articles

1. A Alexander & W Braniff, 'Marginalizing Violent Extremism Online' Lawfare Blog, 21st January 2018, available at https://www.lawfareblog.com/marginalizing-violent-extremism-online accessed on 15 December 2020
2. A Cháves, 'How QAnon Conspiracy Theories Spread in my Colorado Hometown' The Intercept, 23 September 2020, available at https://theintercept.com/2020/09/23/qanon-conspiracy-theory-colorado/ accessed on 15 December 2020
3. A Cichocka, 'To counter conspiracy theories, boost well-being' Nature 587, 177 (12 November 2020) available at https://www.nature.com/articles/d41586-020-03130-6 accessed on 16 December 2020
4. A Hutchinson, 'Facebook announces ban on anti-vax ads, launches new push to promote vaccine use' Social Media Today, 13 October 2020, https://www.socialmediatoday.com/news/facebook-announces-ban-on-anti-vax-ads-launches-new-push-to-promote-vaccin/586939/ accessed 16 December 2020
5. A LaFrance, 'QAnon: Nothing Can Stop What is Coming' The Atlantic, June 2020, available at https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/ accessed 15 December 2020
6. A LaFrance, 'The Prophecies of Q' The Atlantic (2020) accessed at https://www.theatlantic.com/magazine/archive/2020/06/qanon-nothing-can-stop-what-is-coming/610567/ on 17 January 2021

7. A Romano, 'Kicking people off social media isn't about free speech' Vox, January 21 2021, available at https://www.vox.com/culture/22230847/deplatforming-free-speech-controversy-trump accessed February 12 2021

8. Arthur, 'Facebook emotion study breached ethical guidelines, researchers say' The Guardian, 30 June 2014 accessed February 12 2021

9. Associated Press, 'Facebook shuts accounts in anti-vaccine influencer campaign' The Guardian, Wednesday 11 August 2021 https://www.theguardian.com/world/2021/aug/11/facebook-shuts-accounts-in-anti-vaccine-influencer-campaign-fazze?CMP=fb_gu&utm_medium=Social&utm_source=Facebook&fbclid=IwAR0busBMW9n5o9s62ykeJIIRtqmOyolVuGxAF2OHPRvWi4T9iSeBTbxybhg#Echobox=1628672694 accessed 14 August 2021

10. B Bollen, 'Neoliberal Technology is Shaping our World. Let's Turn the Tables on It' Compass Online, 5 January 2017, available at https://www.compassonline.org.uk/neoliberal-technology-is-shaping-our-world-lets-turn-the-tables-on-it/ accessed 10 December 2020

11. B Clifford & H Powell, 'De-platforming and the Online Extremist's Dilemma' Lawfare Blog, 6th June 2019, available at https://www.lawfareblog.com/de-platforming-and-online-extremists-dilemma accessed January 17 2021

12. B Zadrozny & B Collins, 'How Three Conspiracy Theorists Took Q and sparked QAnon' NBC News, 14 August 2018, available at https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-qanon-n900531 accessed 16 December 2020

13. Beckett, 'Facebook bans extremist boogaloo group from its platforms' 30 June 2020, The Guardian, https://www.theguardian.com/technology/2020/jun/30/boogaloo-facebook-ban-extremist-group accessed 17 January 2021

14. C Criddle, 'Facebook sued over Cambridge Analytica Data Scandal' BBC News (28 October 2020) available at https://www.bbc.co.uk/news/technology-54722362 accessed 10 January 2021

15. C Edmond, 'Conspiracy theories have flourished during the pandemic – here's how to stop them in their tracks,' 02 Sept 2020, World Economic Forum https://www.weforum.org/agenda/2020/09/conspiracy-theories-prevent-spread-covid-19-unesco/ accessed 17 December 2021

16. Calamur, 'what is the internet research agency' (2018) The Atlantic Feb 16 2018 https://www.theatlantic.com/international/archive/2018/02/russia-troll-farm/553616/ accessed 10 June 2021

17. Centre for Human Technology, 'Policy Principles' available at https://www.humanetech.com/policy-principles accessed 10 June 2021

18. Corbyn, 'Facebook experiment boosts US voter turnout' (2012) Nature https://doi.org/10.1038/nature.2012.11401 accessed 18 July 2021

19. D Hudson Jr, 'Counterspeech Doctrine' The First Amendment Encyclopaedia, updated December 2017, available at https://www.mtsu.edu/first-amendment/article/940/counterspeech-doctrine accessed 14 March 2021

20. Davidow, 'Skinner Marketing: We're the Rats, and Facebook Likes are the Reward' (2013) The Atlantic https://www.theatlantic.com/technology/archive/2013/06/skinner-marketing-were-the-rats-and-facebook-likes-are-the-reward/276613/ accessed 10 August 2021

21. Dyakon, 'poynter's mediawise training significantly increases people's ability to detect disinformation, new Stanford study finds' December 14, 2020, Poynter, https://www.poynter.org/news-release/2020/poynters-mediawise-training-significantly-increases-peoples-ability-to-detect-disinformation-new-stanford-study-finds/ accessed 10 August 2021 accessed 10 August 2021

22. E Ellis, 'The Influencer Scientists Debunking Online Misinformation' Wired, 13th November 2019, available at https://www.wired.com/story/youtube-misinformation-scientists/ accessed 10 December 2020

23. E Macaskill & G Dance, 'NSA Files: Decoded: What the revelations made for you' The Guardian (1 November 2013) available at https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded accessed 13 August 2021

24. E Salam, 'Majority of Covid misinformation came from 12 people, report finds' The Guardian, 17 July 2021, available at https://www.theguardian.com/world/2021/jul/17/covid-misinformation-conspiracy-theories-ccdh-report accessed 2 September 2021

25. Evans, 'Anti-Semitism Lurks Behind Modern Conspiracy Theories' The Irish Times, 16 February 2021 available at https://www.irishtimes.com/opinion/anti-semitism-lurks-behind-modern-conspiracy-theories-1.4485495 accessed 14 March 2021

26. F Lewsey, 'Cambridge Game 'Pre-Bunks' Coronavirus Conspiracies' accessed at https://www.cam.ac.uk/stories/goviral accessed 12 December 2020

27. Facebook, 'Taking action against people who repeatedly share misinformation' Facebook Newsroom, 26 May 2021, available at https://about.fb.com/news/2021/05/taking-action-against-people-who-repeatedly-share-misinformation/ accessed 10 August 2021

28. Funke & Benkelman, 'want to be a better fact-checker? Play a game' Poynter July 18 2019 https://www.poynter.org/fact-checking/2019/want-to-be-a-better-fact-checker-play-a-game/ accessed 10 December 2020

29. G Benjamin, 'Gaps in UK regulation of online platforms make it difficult to tackle systemic issues – here are some ways we can fix this' LSE Blogs, 23 September 2020, available at https://blogs.lse.ac.uk/medialse/2020/09/23/gaps-in-uk-regulation-of-online-platforms-make-it-difficult-to-tackle-systemic-issues-here-are-some-ways-we-can-fix-this/ accessed 10 December 2020

30. G Smith, 'A Ten Point Rule of Law Test for a Social Media Duty of Care' CyberLeagle, 16 March 2019, available at https://www.cyberleagle.com/2019/03/a-ten-point-rule-of-law-test-for-social.html accessed 12 March 2021

31. H Allam, 'Right Wing Embrace of Conspiracy is 'Mass Radicalization' Experts Warn' NPR, December 15 2020, available at https://www.npr.org/2020/12/15/946381523/right-wing-embrace-of-conspiracy-is-mass-radicalization-experts-warn accessed 10 January 2021

32. Halliday, 'Twitter's Tony Wang: We are the free speech wing of the free speech party' The Guardian, 22 March 2012, available at https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech accessed 11 July 2021

33. Henley, 'How finland starts its fight against fake news in primary schools' The guardian 29 January 2020 https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news accessed 8 August 2021

34. Horwitz & Seetharaman, 'Facebook Executives shut down efforts to make the site less divisive' The Wall Street Journal, 26th May 2020, https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499?mod=hp_lead_pos5 accessed 5 August 2021

35. I Haimowitz, 'No One is Immune: The Spread of Q-anon Through Social Media and the Pandemic' Centre for Strategic and International Studies, 17th December 2020, available at https://www.csis.org/blogs/technology-policy-blog/no-one-immune-spread-q-anon-through-social-media-and-pandemic accessed 10 January 2021

36. Institute for the Future, 'Building a Healthy Cognitive Immune System: defending democracy in the disinformation age' (2019) available at https://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_ODNI_Cognitive_Immunity_Map__2019.pdf accessed 7 August 2021

37. J Alexander, 'YouTube Bans Stefan Molyneux, David Duke, Richard Spencer, and More For Hate Speech' The Verge, 29 June 2020, available at https://www.theverge.com/2020/6/29/21307303/youtube-bans-molyneux-duke-richard-spencer-conduct-hate-speech accessed 14 August 2021

38. J E Uscinski, 'Conspiracy Theory Vs Free Speech – Should We Regulate Social Media?' Bennett Institute For Public Policy, Cambridge University, 20 March 2019, Available at https://www.bennettinstitute.cam.ac.uk/blog/conspiracy-theory-vs-free-speech-should-we-regulat/ accessed 12 March 2021

39. J Kramer, 'Why People Latch On To Conspiracy Theories, According to Science' National Geographic, 8 January 2021, available at https://www.nationalgeographic.com/science/article/why-people-latch-on-to-conspiracy-theories-according-to-science accessed March 12 2021

40. J Mulhall, 'Deplatforming Works: Let's Get on With It' Hope Not Hate, 4th October 2019, available at https://www.hopenothate.org.uk/2019/10/04/deplatforming-works-lets-get-on-with-it/ accessed 10 January 2021

41. J Nicas, 'Can Youtube Quiet Its Conspiracy Theorists?' The New York Times (2 March 2020) available at https://www.nytimes.com/interactive/2020/03/02/technology/youtube-conspiracy-theory.html accessed 10 July 2021

42. J Ohrvik-Stott & C Miller, 'A Digital Duty of Care: Doteveryone's Perspective' Doteveryone, February 2019, available at https://doteveryone.org.uk/wp-content/uploads/2019/02/Doteveryone-briefing-a-digital-duty-of-care.pdf accessed 10 July 2021

43. J Winter 'Exclusive: FBI document warns conspiracy theories are a new domestic terrorism threat' Yahoo News 1 August 2019 available at https://news.yahoo.com/fbi-documents-conspiracy-theories-terrorism-160000507.html?guccounter=1 accessed 12 August 2021

44. K Anderson, 'How America Lost Its Mind' The Atlantic, 28 December 2017 available at https://www.theatlantic.com/magazine/archive/2017/09/how-america-lost-its-mind/534231/ accessed 3 January 2021

45. K Roose, 'The Making of a YouTube Radical' The New York Times, 8 June 2019, available at https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html accessed 10 July 2021

46. K Starbird, E S Spiro & K Koltai, 'misinformation, crisis, and public health – reviewing the literature (June 25 2020) Social Science Research Council, Media Well, available at http://doi.org/10.35650/MD.2063.d.2020 accessed 17 July 2021

47. Lessenski, 'Just think about it. Findings of the media literacy index 2019' (November 2019) open society institute Sofia, Policy Brief 55 https://osis.bg/wp-content/uploads/2019/11/MediaLiteracyIndex2019_-ENG.pdf accessed 14 March 2021

48. M Bergen, 'YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant" Bloomberg, 2nd April 2019, available at https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant accessed 16 July 2021

49. M Isaac, 'Facebook Says it Will Remove Coronavirus Vaccine Misinformation' December 3 2020, New York Times, accessed at https://www.nytimes.com/2020/12/03/technology/facebook-coronavirus-vaccine-misinformation.html accessed 16 July 2021

50. Mac & Silverman, 'Facebook has been showing military gear ads next to insurrection posts' Buzzfeed News, January 13 2021, available at https://www.buzzfeednews.com/article/ryanmac/facebook-profits-military-gear-ads-capitol-riot accessed 17 August 2021

51. Maheshwari, 'AT&T to advertise on youtube again after a nearly 2-year holdout' The New York Times, 18th January 2019, available at https://www.nytimes.com/2019/01/18/business/media/att-youtube-advertising.html accessed 16 August 2021

52. N Cofnas, 'Deplatforming Wont Work' Quillette, 8th July 2019, available at https://quillette.com/2019/07/08/deplatforming-wont-work/ accessed 16 March 2021

53. National Literary Trust, 'commission on fake news and the teaching of critical literacy skills in schools' (2018) https://literacytrust.org.uk/policy-and-campaigns/all-party-parliamentary-group-literacy/fakenews/ accessed 10 August 2021

54. O Papakyriakopoulos et al, 'The Spread Of COVID-19 Conspiracy Theories On Social Media And The Effect Of Content Moderation' Harvard Kennedy School Misinformation Review, 18 August 2020, Available At https://misinforeview.hks.harvard.edu/article/the-spread-of-covid-19-conspiracy-theories-on-social-media-and-the-effect-of-content-moderation/ accessed 10 August 2021

55. P Ball & A Maxmen, 'The Epic Battle Against Coronavirus Misinformation and Conspiracy Theories' Nature, 27 May 2020, available at https://www.nature.com/articles/d41586-020-01452-z accessed 10 July 2021

56. P Noor, 'Does Suppressing Online Conspiracy Theorists Work? Experts Weigh In' 30 July 2020, The Guardian, accessed at https://www.theguardian.com/media/2020/jul/30/qanon-does-suppressing-online-conspiracy-theorists-work accessed 10 March 2021

57. Policy@Manchester, 'Infodemic: tackling conspiracy theories on social media' Politics Home, 29 march 2021 https://www.politicshome.com/members/article/infodemic-tackling-conspiracy-theories-on-social-media accessed 11 March 2021

58. R Ruiz, 'No one's talking about the complex relationship between disinformation and mental health. That changes now' Mashable, 27 June 2021, available at https://mashable.com/article/mental-health-disinformation-conspiracy-theories-depression accessed 10 August 2021

59. Radicalisation Awareness Network, 'Harmful conspiracy myths and effective P/CVE countermeasures' 02/10/2020, 2 accessed at https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ran-papers/docs/ran_paper_conspiracy-myths-expert-meeting_en.pdf accessed 10 March 2021

60. Read, 'how can we tackle conspiracy theories in schools' tes, 28th march 2021 https://www.tes.com/news/how-can-we-tackle-conspiracy-theories-schools accessed 10 August 2021

61. Reuters Institute 'mainstream media need to make news more accessible, relevant and entertaining to attract next generation of consumers' (University of Oxford) https://reutersinstitute.politics.ox.ac.uk/risj-review/mainstream-media-need-make-news-more-accessible-relevant-and-entertaining-attract-next accessed 12 August 2021

62. Roose & Conger, 'YouTube to Remove Thousands of Videos Pushing Extreme Views' The New York Times, 5th June 2019, available at https://www.nytimes.com/2019/06/05/business/youtube-remove-extremist-videos.html accessed 17 August 2021

63. S Rodriguez, 'here are the scandals and other incidents that have sent Facebook's share price tanking in 2018' CNBC (20 November 2018) available at https://www.cnbc.com/2018/11/20/facebooks-scandals-in-2018-effect-on-stock.html accessed 19 August 2021

64. Shellenbarger, 'Most students don't know when news is fake, Stanford study finds' November 21 2016, The Wall Streat Journal https://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576 accessed 13 August 2021

65. Speller, 'Reducing Harm in Social Media Through a Duty of Care' LSE Blogs, 10th May 2018, available at https://blogs.lse.ac.uk/medialse/2018/05/10/reducing-harm-in-social-media-through-a-duty-of-care/ accessed 10 March 2021

66. Stop Hate For Profit, 'One year after stop hate for profit: progress update from the coalition' Stop Hate for Profit, 30 July 2020 https://www.stophateforprofit.org/   accessed 10 August 2021

67. T Sharot, 'To quell misinformation, use carrots – not just sticks' Nature 591, 347 (2021) accessed at https://www.nature.com/articles/d41586-021-00657-0 accessed 10 March 2021

68. T Simonite, 'Facebook's AI for Hate Speech Improves. How Much Is Unclear' Wired, 5th December 2020, available at https://www.wired.com/story/facebook-ai-hate-speech-improves-unclear/ accessed 10 January 2021

69. The Economist, 'facebook eyes a future beyond social media' The Economist, 29th July 2021, available at https://www.economist.com/business/2021/07/29/facebook-eyes-a-future-beyond-social-media accessed 10 August 2021

70. The YouTube Team, 'Managing Harmful Conspiracy Theories on YouTube' YouTube Official Blog, 15 October 2020, available at https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/ accessed 7 July 2021

71. The YouTube Team, 'Our Ongoing Work to Tackle Hate' YouTube Official Blog, 5th June 2019, available at https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate accessed 7 July 2021

72. TikTok UK, 'NEW! Screen Time Management and Restricted Mode Features on Tiktok' Medium, 4 April 2019, available at https://medium.com/@TikTokUK/new-screen-time-management-and-restricted-mode-features-on-tiktok-86eb30bcf93d accesed 7 July 2021

73. V Heffernan, 'Our best hope for civil discourse online is on…Reddit' Wired, 16th January 2018, available at https://www.wired.com/story/free-speech-issue-reddit-change-my-view/ accessed 13 August 2021

Websites

1. 'Bad News' Game accessed at https://www.getbadnews.com/#intro accessed 10 January 2021
2. 'Go Viral' Game accessed at https://www.goviralgame.com/en accessed 10 January 2021

Podcasts

1. Your Undivided Attention, 'The Cure for Hate' 19 December 2019 accessed https://your-undivided-attention.simplecast.com/episodes/the-cure-for-hate-FOjkuIcA accessed 17 August 2021
2. The New York Times, 'Rabbit Hole' Podcast Series, 17 April 2020, accessed at https://www.nytco.com/press/introducing-rabbit-hole-a-new-narrative-audio-series-from-the-new-york-times/ accessed 17 July 2021

Videos and TV

1. E Wang, 'Why Some Conspiracy Theories Just Won't Die' 26th March 2021, TEDxDuke available at https://www.ted.com/talks/elise_wang_why_some_conspiracy_theories_just_won_t_die/up-next accessed 10 August 2021
2. J Orlowski, 'The Social Dilemma' Netflix, 27th August 2020, available at https://www.netflix.com/gb/title/81254224 accessed 10 December 2020