DEMOCRACY AND HUMAN RIGHTS

CANARIES IN THE COALMINE

Building Resilience to Online Misinformation among Young People in the UK

Daniel Allington, Harry Shukman, William Galinsky and Rachel Briscoe
January 2023

→ The spread of online misinformation is a key concern in a digitally-permeated society, prompting fears that engagement with it can lead, particularly among young people, to radicalisation and inter-group hate.

→ This paper assesses the scale of the problem, and examines a range of tools and techniques intended to mitigate its impact, from online fact-checking services and content warnings to interfaith dialogue and state-mandated educational programmes.

→ The most promising methods of confronting misinformation are likely those that engage people in the real-world and build lasting relationships across community divides beyond the internet. Interventions that take place solely online, on the other hand, are yet to deliver robust and easily scalable impacts.

FRIEDRICH EBERT STIFTUNG

CANARIES IN THE COALMINE

Building Resilience to Online Misinformation among Young People in the UK

Contents

Jewish communities are deeply troubled by the promotion of online antisemitic hatred and abuse. JW3 welcomed the opportunity to work with FES and commissioned a respected academic, Dr Daniel Allington, and a noted investigative journalist, Harry Shukman, with experience in this area, to present their own assessment and views of the dangers, as well as potential online responses. As a follow-up to their work, JW3 is exploring how it is uniquely placed as a Jewish Cultural and Community centre to develop some of the novel interventions described in the conclusion of this paper, including partnering with interactive theatre creators Fast Familiar in co-producing an experience that responds to these and other challenges by working at a local level through arts and culture. This paper represents the views of the authors and not necessarily the views of JW3.

INTRODUCTION

The creators of the long-running British soap opera *Coronation Street* recently unveiled a storyline intended to explore (in the words of the show's producer) the »very 21st century problem« of »teenagers self-radicalising through watching extreme content online«.[1] The idea of such a problem was not the product of a dramatist's imagination: the chief of MI5, the British Security Service, has also raised concerns about children being exposed to what he called »a confused soup of hate« within which some were being influenced towards extreme views.[2] But how much of a problem is this really? And what can be done about it?

In the UK, the word »disinformation« is used to designate the »deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences«, while the word »misinformation« conversely »refers to the inadvertent sharing of false information«.[3] A 2019 report by the UK Commission for Countering Extremism warned that »extremists actively spread disinformation to shape public attitudes in line with their ideological worldview« and that »[e]xtremism thrives in an environment where alternative sources of news promote hateful worldviews and denigrate the mainstream media«.[4] Al-

though it appears likely that the Internet's increasing role as a vector for radicalisation is in line with an increase in »online activity by society generally«,[5] and thus that there is no specific harm associated with the Internet *per se*, this does not mean that there is no need to find solutions to the problem of hate, extremism, and misinformation that are tailored to the digitally-permeated world in which young people especially now live. That is, just as some popular social platforms may have been inadvertently designed in a way that benefits the distribution of misinformation,[6] it may make sense to leverage the affordances of the Internet in ways that benefit inter-group harmony and the reinforcement of reliable information.

This paper focuses on misinformation, hate, and radicalisation in relation to the online environment, with a particular focus on young people. It surveys the research that has been done on the problems that exist, and also takes a critical look at interventions that have been attempted in the interests of limiting the spread of misinformation, countering hatred between groups, and preventing people from being radicalised. We then consider a possible role for community and cultural organisations in developing more nuanced interventions.

MISINFORMATION, RADICALISATION, AND THE INTERNET

Some of the best-studied examples of harmful online content concern vaccination and COVID-19. A study comparing data collected in two UK and two US surveys found hesitancy with regard to COVID-19 vaccination to be associated with non-use of traditional media as a source of information in both countries, although the evidence with regard to youth (which was found to have an association with vaccine hesitancy in both UK surveys, but only in one of the US surveys) and social media use (which was found to have an association with vaccine hesitancy in one UK study only) was more mixed,[7] and one further study found a very small negative association between vaccine hesitancy and age in the US.[8] Literature reviews of studies in many different national contexts have found a correlation between belief in misinformation about COVID-19 (such as conspiracy theories regarding the origin of the virus) and reluctance to be vacci-

1 Quoted in Molly Moss, »Coronation Street Confirms Max Turner to Be Groomed by Extremists: The Storyline Will Examine How Extremist Groups Target, Groom and Operate«, *Radio Times*, 27 October 2022, https://www.radiotimes.com/tv/soaps/coronation-street/coronation-street-max-turner-grooming-storyline-newsupdate/.

2 Quoted in Dan Sabbagh, »MI5 Chief: UK Will Have to Tackle Russian Aggression for Years to Come«, *Guardian*, 16 November 2022, https://www.theguardian.com/uk-news/2022/nov/16/mi5-chief-uk-will-have-to-tackle-russian-aggression-for-years-to-come.

3 UK Government, »Disinformation and ›Fake News‹: Interim Report: Government Response to the Committee's Fifth Report« (London: UK Government, 2018), https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1630/163002.htm.

4 CCE, »Challenging Hateful Extremism« (London: Commission for Countering Extremism, 7 October 2019), 38, https://www.gov.uk/government/publications/challenging-hateful-extremism.

5 Jonathan Kenyon, Jens Binder, and Christopher Baker-Beall, »The Internet and Radicalisation Pathways: Technological Advances, Relevance of Mental Health and Role of Attackers« (Ministry of Justice, 2022), 2, https://www.gov.uk/government/publications/internet-and-radicalisation-pathways-technological-advances-relevance-of-mental-health-and-role-of-attackers.

6 E.g. Daniel Allington and Tanvi Joshi, »›What Others Dare Not Say‹: An Antisemitic Conspiracy Fantasy and Its YouTube Audience«, *Journal of Contemporary Antisemitism* 3, no. 1 (2020): 49–50, https://doi.org/10.26613/jca/3.1.42.

7 Daniel Allington et al., »Media Usage Predicts Intention to Be Vaccinated against SARS-CoV-2 in the US and the UK«, *Vaccine* 39 (2021): 2595–2603, https://doi.org/10.1016/j.vaccine.2021.02.054.

8 Timothy Callaghan et al., »Correlates and Disparities of COVID-19 Vaccine Hesitancy«, *Social Science & Medicine* 272 (2021): 113638.

nated against COVID-19,[9] as well as between social media use and reluctance to be vaccinated.[10] This suggests that greater use of social media among younger people in developed countries might be associated with greater vaccine hesitancy, perhaps as a result of greater exposure to misinformation. Indeed, studies have found that young people in the UK are less likely to take up the opportunity to be vaccinated against COVID-19, even after controlling for such variables as ethnicity, education, experiences of NHS healthcare, and trust in government, scientists, and medics.[11] One further study carried out in Ireland and the UK found that while youth was only associated with vaccine resistance in the UK, »vaccine resistant individuals in both countries were less likely to consume, and trust, information from ›traditional‹ sources (i.e. newspapers, television, radio, and government agencies), and were somewhat more likely to obtain information from social media channels«.[12] COVID-19 misinformation has also been linked to vandalism of mobile phone masts and attacks on telecoms workers (apparently thanks to a false belief that COVID-19 symptoms are caused not by a virus but by 5G radiation),[13] as well as to increased hate crimes against South and East Asians (who were wrongly associated with the spread of the pandemic).[14]

While COVID-19 misinformation has only been linked to property damage and relatively low-level violence, the myth of »White Genocide« or the »Great Replacement« appears by contrast to have been a motive for a number of mass shooters and apparent attempted mass shooters. This conspiracy theory alleges that immigration from non-white, predominantly Muslim populations into majority-white, historically Christian nations is intentionally promoted, whether by Jews or by »liberal elites«, as a way of weakening the majority populations of these nations.[15] The idea of the »Great Replacement« is usually attributed to Renaud Camus, but it can be traced back to Adolf Hitler.[16] Terrorists motivated by belief in the »Great Replacement« include the Pittsburgh synagogue shooter, the Christchurch mosque shooter, the El Paso mall shooter, the Poway synagogue shooter, the Oslo mosque shooter, and the Halle synagogue shooter, as well as Anders Breivik, the Norwegian domestic terrorist who killed 77.[17] Like all mass killings, these are unique incidents, but a series of studies carried out in Norway and Denmark found support for the view that belief in the »Great Replacement« drives increased Islamophobia and is associated with violent intentions towards Muslims.[18] However, as the list of terror attacks in the previous sentence makes clear, other groups held responsible for Muslim immigration have also been targeted by individuals who took this conspiracy theory to be accurate – including liberals (Breivik's target) and Jews (the intended target of the various synagogue attacks).

Although many proponents of the »Great Replacement« theory, including Renaud Camus, show no inclination towards antisemitism, it forms part of the antisemitic mythology built around the conspiratorial idea of the so-called »Zionist Occupation Government« or ZOG, which has long been a key plank within the ideology of right-wing extremist groups. As one extremism researcher writes,

> the [concept of the] »Zionist Occupation Government« … is based on the idea that … most … white countries are occupied and controlled by Jews and their lackeys. This ideology, developed by US racist movements such as Aryan Nations … ha[s] been … developed into a highly violent revolutionary and terrorist doctrine. Every element of the Establishment – politicians, the bureaucracy, the police, the military, the media, and the intelligentsia – is now regarded as the lackey and tool of ZOG. … The only way to save the pure Aryan race from extermination is [argued to be] by fighting a total racial war against the »Zionist Occupation Government« and its lackeys.[19]

9   Ingjerd Skafle et al., »Misinformation about COVID-19 Vaccines on Social Media: Rapid Review«, *Journal of Medical Internet Research* 24, no. 8 (2022): e37367, https://doi.org/10.2196/37367.

10 Fidelia Cascini et al., »Social Media and Attitudes towards a COVID-19 Vaccination: A Systematic Review of the Literature«, *EClinicalMedicine* 48 (2022): 101454, https://doi.org/10.1016/j.eclinm.2022.101454.

11 Daniel Allington et al., »Coronavirus Conspiracy Suspicions, General Vaccine Attitudes, Trust, and Coronavirus Information Source as Predictors of Vaccine Hesitancy among UK Residents during the COVID-19 Pandemic«, *Psychological Medicine*, 2021, https://doi.org/10.1017/S0033291721001434; Daniel Allington et al., »Trust and Experiences of National Health Service Healthcare Do Not Fully Explain Demographic Disparities in Coronavirus Vaccination Uptake in the UK: A Cross-sectional Study«, *BMJ Open* 12, no. e053827 (2022), https://doi.org/10.1136/ bmjopen-2021-053827.

12 Jamie Murphy et al., »Psychological Characteristics Associated with COVID-19 Vaccine Hesitancy and Resistance in Ireland and the United Kingdom«, *Nature Communications* 12, no. 29 (5 June 2020), https://doi.org/10.1038/s41467-020-20226-9, p. 10.

13 Jim Waterson and Alex Hern, »At Least 20 UK Phone Masts Vandalised over False 5G Coronavirus Claims: Industry Body Assures People in Open Letter There Is No Link between 5G and Pandemic«, *The Guardian*, 6 April 2020, https://www.theguardian.com/technology/2020/apr/06/at-least-20-uk-phone-masts-vandalised-over-false-5g-coronavirus-claims; Alex Hern, »5G Conspiracy Theories Fuel Attacks on Telecoms Workers: One Openreach Engineer Reports Being Forced to Flee Angry Group since Covid-19 Crisis«, *The Guardian*, 7 May 2020, https://www.theguardian.com/business/2020/may/07/5g-conspiracy-theories-attacks-telecoms-covid.

14 Jamie Grierson, »Anti-Asian Hate Crimes up 21% in UK during Coronavirus Crisis: Ministers Also Tell Commons Committee Revenge Porn and Online Fraud Have Increased«, *The Guardian*, 13 May 2020, https://www.theguardian.com/world/2020/may/13/anti-asian-hate-crimes-up-21-in-uk-during-coronavirus-crisis.

15 Tina Askanius, »On Frogs, Monkeys, and Execution Memes: Exploring the Humor-Hate Nexus at the Intersection of Neo-Nazi and Alt-Right Movements in Sweden«, *Television and New Media* 22, no. 2 (2021): 157, https://doi.org/10.1177/1527476420982234; Mattias Ekman, »The Great Replacement: Strategic Mainstreaming of Far-Right Conspiracy Claims«, *Convergence: The International Journal of Research into New Media Technologies* 28, no. 4 (2022): 1311, https://doi.org/10.1177/13548565221091983.

16 Allington and Joshi, »›What Others Dare Not Say‹: An Antisemitic Conspiracy Fantasy and Its YouTube Audience«, 38.

17 Allington and Joshi, 36–37.

18 Milan Obaidi et al., »The ›Great Replacement‹ Conspiracy: How the Perceived Ousting of Whites Can Evoke Violent Extremism and Islamophobia«, *Group Processes and Intergroup Relations* 25, no. 7 (2022): 1675–95, https://doi.org/10.1177/13684302211028293.

19 Tore Bjørgo, »Militant Neo-Nazism in Sweden«, *Terrorism and Political Violence* 5, no. 3 (1993): 36–37, https://doi.org/10.1080/09546559308427219.

The use of the term »Zionist« in the above may come as a surprise to those who are familiar with antizionism only as a »leftist ticket«.[20] However, placing the focus on »Zionists« rather than »Jews« as conspirators began as a right-wing extremist tactic to avoid accusations of antisemitism, and was noted as long ago as 1970.[21] Over time, the notion of a global »Zionist« conspiracy came to be adopted by many across the far right, the far left, and the Islamist movement, such that medieval European antisemitic ideas have become legitimated for new generations – as in the discourse of the young British Muslim who told an interviewer that »Israelis are a cruel … an evil group of people« who »just want to get rich quick« and were »[h]istorically … involved in murdering kids and innocent people« and yet still insisted that hating Israelis had »nothing to do with« hating Jews.[22] One noted antisemitism researcher explains the phenomenon as follows:

> In our day, the word »Zionism« … often functions as an antisemitic curse word, which hurts and discredits Jews. … The word »Zionism« frequently constructs Jews as participating in dishonest global networks, conspiracies of lies and propaganda, in their own selfish interest. It positions most Jews as though they are in alliance with a formidable global system of oppression, sometimes called »modernity«, or »capitalism« or »imperialism«.[23]

This positioning is often accomplished by ignoring the Holocaust as a reason for Israel's existence, and thus presenting Israel as a »US-backed settler colony« or even as »an outstanding perpetrator state« and »spearhead« of the »neocolonial capitalist world system«,[24] in an unironic echo of the more obviously antisemitic early 20th century positioning of Jews as colonisers of Europe.[25] It is also at times accomplished by equating Israel's actions with those of the Nazis, and thereby »erasing the determinate role of exterminatory antisemitism in the Holocaust«,[26] or through more elaborate theoretical means, as in the academic argument that Israel represents a »historical continuity of hetero-conquest whereby colonial conquest dictates a violent process of racialised, gendered, and sexualised classifications that infuse its politics of Time and Space«.[27] It becomes hard to see such abstractions as anything more than a fig leaf, however, when one considers

such a phenomenon as the convoy of pro-Palestinian activists who chose to respond to the 2021 Israel-Hamas conflict by driving through North London with loudspeakers broadcasting the message »fuck the Jews, rape their daughters«.[28]

What the above shows is that a considerable problem exists with regard to the circulation and endorsement of misinformation – including misinformation that primarily threatens the individual who endorses it (e.g. incorrect beliefs about vaccination) and hateful misinformation about members of different social groups. But while all this is highly concerning, it is important not to be alarmist: although it appears that there is a notable minority of young people attracted to misinformation and extremist online content, this does not mean that an army of young terrorists is forming. A quantitative study of people convicted of terror offences in England and Wales compared those who primarily radicalised online with those who primarily radicalised offline and those whose radicalisation process involved both online and offline components. It found that members of the former category were different from members of the latter two, both in terms of their level of integration with extremist groups and in terms of their level of involvement in terrorist violence. As the authors write,

> Those who primarily radicalised online were less likely to be socially connected in the context of the offence and they were more likely to display signs of mental illness or personality disorder …

> Most importantly, differences were found in assessed levels of engagement, intent, and capability, with those who primarily radicalised online considered the least identified with an extremist group or cause, and least willing and able to perpetrate violent extremist acts.[29]

This means that, while individuals can certainly reach the point of committing terror offences as a result of exposure to extremist online content, they do not necessarily join extremist groups or take steps towards the perpetration of violence.[30] Moreover, a subsequent iteration of the same research project not only confirmed the above-quoted findings, but also established that, where terror convicts were

20 Julia Edthofer, »Israel as Neo-Colonial Signifier? Challenging de-Colonial Anti-Zionism«, *Journal for the Study of Antisemitism* 7 (2015): 31.

21 Patterns of Prejudice, »›Anti-Zionists‹ and Antisemites«, *Patterns of Prejudice* 4, no. 4 (1970): 28.

22 Rusi Jaspal, *Antisemitism and Anti-Zionism: Representation, Cognition, and Everyday Talk* (Abingdon: Routledge, 2014), 168.

23 David Hirsh, »How the Word ›Zionist‹ Functions in Antisemitic Vocabulary«, *Journal of Contemporary Antisemitism* 4, no. 2 (2021): 3.

24 Edthofer, »Israel as Neo-Colonial Signifier? Challenging de-Colonial Anti-Zionism«, 32.

25 See Tony Barta, »Discourses of Genocide in Germany and Australia: A Linked History«, *Aboriginal History* 25 (2001): 38.

26 Matt Bolton, »Conceptual Vandalism, Historical Distortion: The Labour Antisemitism Crisis and the Limits of Class Instrumentalism«, *Journal of Contemporary Antisemitism* 3, no. 2 (2020): 24, https://doi.org/10.26613/jca/3.2.56.

27 Walaa Alqaisya, »Palestine and the Will to Theorise Decolonial Queering«, *Middle East Critique* 29, no. 1 (2020): 104, https://doi.org/10.1080/19436149.2020.1704505.

28 Gabriella Swerling, »Men Arrested over Pro-Palestinian Convoy Linked to Previous Anti-Semitic Incident: It Is Alleged That Two Vehicles Followed a Jewish Man in Salford before the Passengers Exited Their Cars and Attempted to Smash His Windows«, *Daily Telegraph*, 18 May 2021, https://www.telegraph.co.uk/news/2021/05/18/.men-arrested-pro-palestine-london-convoy-linked-previous-anti/; BBC, »Pro-Palestine Rally: Racial Hatred Charges Dropped for Two More Men«, *BBC Local News*, 21 November 2022, https://www.bbc.co.uk/news/uk-england-london-63707451.

29 Jonathan Kenyon, Jens Binder, and Christopher Baker-Beall, »Exploring the Role of the Internet in Radicalisation and Offending of Convicted Extremists« (London: Ministry of Justice, 2021), 2, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1017413/exploring-role-internet-radicalisation.pdf.

30 It is possible for someone convicted of a terrorist offence not to have engaged in terrorism *per se*: in the UK, for example, it is an offence to assemble, possess, or view a document or other record containing »information of a kind likely to be useful to a person committing or preparing an act of terrorism«; Terrorism Act (2000), sec. 58 (1), https://www.legislation.gov.uk/ukpga/2000/11/section/58.

involved in violent plots, the attacks had been successful in just 18 per cent - just one of the 17 cases involving individuals radicalised online, as compared to 58 per cent of the 55 cases involving individuals radicalised offline.[31] In this connection, it is important to take note of a University of Oslo study which appears to suggest that young right-wing extremists are more likely to be »keyboard warriors« than terrorists as conventionally understood:

> Overall, if we look at the age profiles of those [right-wing extremists] who were behind the 26 fatal attacks recorded in Western Europe since 2015, only one perpetrator was below 20 years [old], [and] only five below 25, whereas 12 perpetrators were over 40 years old. In other words, while young extremists may be posing as potential terrorists online, those who do carry out fatal attacks tend to be much older … at least in Western Europe.[32]

As the UK's Independent Reviewer of Terrorism Legislation has observed, at least with regard to young people whose extremist activities are confined to the Internet, we are far less likely to be dealing with hardened terrorists than with often troubled individuals »whose ideological commitment, ripped from the Internet, is skin-deep or a passing fancy or adolescent reaction to the world«.[33] This view appears well grounded in the evidence: the aforementioned study found that just 7% of terror convicts who radicalised online were judged to have »significant« capability to carry out attacks, as opposed to 23% and 38% of the partially or wholly offline convicts.[34] However, it does not imply that we can ignore the problem of young people engaging with extremist content and misinformation online: as the same Independent Reviewer has also argued, while many of those who engage in extremist activities online are simply »showing off, without any intention to commit real world violence«, those who have carried out recent attacks »were [also] once mere keyboard warriors«.[35]

By helping young people to turn away from extremist online content and misinformation at a point before they have truly committed to a cause – for example by joining groups capable of cementing their ideological convictions with real-world social connections and encouraging and materially supporting their fantasies of violence – it may be possible to avert future terror attacks. But how can this be done?

INTERVENTIONS WITH REGARD TO MISINFORMATION

One of the most straightforward interventions with regard to misinformation is that of »fact-checking«. Fact-checking websites like PolitiFact and Snopes aim to verify public statements and gauge their accuracy. While research has shown that fact-checking interventions can reduce belief in misinformation in laboratory settings, doubts remain about the extent of their efficacy in the real world.[36][37] Belief in misinformation can prove remarkably sticky, even when corrected immediately after a reader has come into contact with it.[38] There are also questions of cost. Identifying misinformation, correcting it and effectively disseminating a counter-narrative is time-consuming and requires epistemological expertise.[39] Producing fake news, on the other hand, can be much cheaper as far as resources and time are concerned.[40] When readers encounter fact-checked information, they may respond more favourably if they see it shared by trusted sources such as friends and followers.[41] However, if fact-checking services are perceived to have a partisan agenda, they may have the potential to backfire and inadvertently increase belief in an item of fake news.[42]

Closely related to fact-checking is the practice of adding »content warnings«. Social media companies place these on shared content that they have judged to contain misinformation in the hope that readers will ignore or stop sharing them. The issuing of content warnings appears to reduce the perceived accuracy of fake news, although sometimes the impact can be modest.[43][44] However, social media

31 Kenyon, Binder, and Baker-Beall, »The Internet and Radicalisation Pathways: Technological Advances, Relevance of Mental Health and Role of Attackers«, 61.

32 Anders Ravik Jupskås et al., »Five Myths of Far-Right Violence: Improved Systematic Data on Severe Far-Right Violence in Western Europe Debunks Five Common Myths in the Field«, *Center for Research on Extremism* (blog), 23 November 2022, https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2022/five-myths-of-far-right-violence.html.

33 Jonathan Hall, »Keyboard Warriors or International Terrorists?« (London: Independent Reviewer of Terrorism Legislation, 14 July 2022), 8, https://terrorismlegislationreviewer.independent.gov.uk/wp-content/uploads/2022/07/IRTL-Keyboard.pdf.

34 Kenyon, Binder, and Baker-Beall, »The Internet and Radicalisation Pathways: Technological Advances, Relevance of Mental Health and Role of Attackers«, 63.

35 Hall, »Keyboard Warriors or International Terrorists?«, 2.

36 Ethan Porter and Thomas J. Wood. »The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdo«. PNAS Vol 118, No. 37, September 2021. https://doi.org/10.1073/pnas.2104235118.

37 Jingwen Zhangab, Jieyu Ding Featherstone, Christopher Calabresea, Magdalena Wojcieszaka. »Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines«, Preventive Medicine, 145, 2021. https://doi.org/10.1016/j.ypmed.2020.106408.

38 Emily Thorson. »Belief Echoes: The Persistent Effects of Corrected Misinformation«, Political Communication Volume 33, 2016 - Issue 3, https://doi.org/10.1080/10584609.2015.1102187.

39 Hyunuk Kim and Dylan Walker. »Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media«, Harvard Kennedy School Misinformation Review, 2020, https://doi.org/10.37016/mr-2020-021.

40 European Commission. »Report on the implementation of the Action Plan Against Disinformation«, 2019, https://op.europa.eu/en/publication-detail/-/publication/8a94fd8f-8e92-11e9-9369-01aa75ed71a1/language-en.

41 Drew B. Margolin, Aniko Hannak and Ingmar Weber. »Political Fact-Checking on Twitter: When Do Corrections Have an Effect?«, Political Communication, 35:2, 2018, https://doi.org/10.1080/10584609.2017.1334018.

42 Dino P. Christenson, Sarah E. Kreps, and Douglas L. Kriner. »Contemporary Presidency: Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion«, Presidential Studies Quarterly Vol. 0, No. 0, November 2020, 1–15, DOI: 10.1111/psq.12687.

43 Jack Nassetta and Kimberly Gross. »State media warning labels can counteract the effects of foreign misinformation«, Harvard Kennedy School Misinformation Review, 2020, https://doi.org/10.37016/mr-2020-45.

44 Katherine Clayton et al. »Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media.« Political Behavior, 42:4, 2020, https://doi.org/10.1007/s11109-019-09533-0.

companies are unable to label every item of misinformation on their platforms, leading to the »implied truth effect, where readers feel confident in sharing misinformation that has yet to be labelled under the mistaken assumption that it has been reviewed and deemed acceptable by fact-checkers«.[45][46] In addition, repeated exposure to misinformation – whether it is labelled with a content warning or not – can increase the likelihood that someone believes it.[47] Of the two main types of content warning on social media, interstitial covers that obscure misinformation and require an additional click to view appear to be more effective than contextual tags which do not compel action when it comes to reducing a reader's perceived accuracy of a social media post.[48]

A lesser-known and less-frequently practised intervention is that of cognitive inoculation, or »prebunking«. This is based on the idea that by pre-emptively exposing people to small doses of fake news they can be immunised against misinformation.[49] Cognitive inoculation is thus unlikely to work on an individual who has already encountered and believes in a particular falsehood.[50] This poses a challenge: in hyper-online society, how can people be immunised before they come into contact with misinformation?[51] In response to these issues, researchers have suggested that cognitive inoculation is most effective not when focusing on specific items of fake news but in explaining the techniques used by producers of misinformation.[52,53] Instead of prebunking a vaccine conspiracy theory, for instance, interventions could explain how fake news might impersonate official sources or use heightened emotional language. This technique has

proved promising in experiments.[54] Still, the issue remains that, as with vaccines, any immunity provided by cognitive inoculation seems to decay over time.[55]

A lack of competence in navigating the Internet – or poor digital literacy – is said to correlate with a greater tendency to believe misinformation.[56] Digital literacy interventions can improve the ability to identify misinformation,[57] although some researchers have not found this to be the case.[58] Where digital literacy interventions do positively impact misinformation belief, there are additional concerns about lessons fading with time.[59] As with fact-checking, interventions in digital literacy have the potential to backfire, perhaps by fomenting an unhealthy scepticism of genuine news sources.[60] Significantly, increased digital literacy does not necessarily reduce the likelihood that people will share misinformation.[61] Moreover, Internet users may repost fake news on social media – even if they suspect it is false – to advance political goals.[62]

Lastly, responding to misinformation with an alternative narrative – instead of content moderation or censorship – is known as »counterspeech«. Counterspeech is championed by grassroots campaigns that seek out and respond to misinformation, as well as social media platforms which fund trusted content creators to address media falsehoods. However, there is no guarantee that Internet users at risk of believing in misinformation will come across the videos of those

45 Gordon Pennycook, Adam Bear, Evan T. Collins, David G. Rand. »The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings«, Management Science, 66:11, 2020, https://doi.org/10.1287/mnsc.2019.3478.

46 Katherine Clayton et al., 2020.

47 Gordon Pennycook, Tyrone D. Cannon, and David G. Rand. »Prior exposure increases perceived accuracy of fake news«, Journal of Experimental Psychology, 147:12, 2018, doi: 10.1037/xge0000465.

48 Filipo Sharevski, Raniem Alsaadi, Peter Jachim, and Emma Pieroni. »Misinformation warnings: Twitter's soft moderation effects on COVID-19 vaccine belief echoes«, Computer Security, 2022, March:114, DOI: 10.1016/j.cose.2021.102577

49 Jon Roozenbeek, Sander van der Linden, and Thomas Nygren. »Prebunking interventions based on ›inoculation‹ theory can reduce susceptibility to misinformation across cultures«, The Harvard Kennedy School Misinformation Review, 1:2, 2020. https://doi.org/10.37016//mr-2020-008.

50 Daniel Jolley and Karen M. Douglas. »Prevention is better than cure: Addressing anti-vaccine conspiracy theories«, Journal of Applied Social Psychology, 47:8, 2017, https://doi.org/10.1111/jasp.12453.

51 John Cook, Stephan Lewandowsky, and Ullrich K. H. Ecker. »Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence«. PLoS ONE, 12:5, 2017. https://doi.org/10.1371/journal.pone.0175799.

52 Jon Roozenbeek and Sander van der Linden. »Fake news game confers psychological resistance against online misinformation«, Palgrave Communications, 5:65, 2019. https://doi.org/10.1057/s41599-019-0279-9.

53 Ananya Iyengar, Poorvi Gupta, Nidhi Priya. »Inoculation against conspiracy theories: A consumer side approach to India's fake news problem«, Applied Cognitive Psychology, Early View, 29 August 2022. https://doi.org/10.1002/acp.3995.

54 Melisa Basol, Jon Roozenbeek, and Sander van der Linden. »Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News«, Journal of Cognition, 1:1, 2020. http://doi.org/10.5334/joc.91.

55 John A. Banas & Stephen A. Rains. »A Meta-Analysis of Research on Inoculation Theory«, Communication Monographs, 77:3, http://dx.doi.org/10.1080/03637751003758193.

56 Andrew M. Guess, Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. »A digital media literacy intervention increases discernment between mainstream and false news in the United States and India«, PNAS 117:27, 2020, https://doi.org/10.1073/pnas.1920498117.

57 Thomas Nygren, Mona Guath, Carl-Anton Werner Axelsson, and Divina Frau-Meigs. »Combatting Visual Fake News with a Professional Fact-Checking Tool in Education in France, Romania, Spain and Sweden«, Information, 12:5, 2021, https://doi.org/10.3390/info12050201.

58 Sumitra Badrinathan, »Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India«, American Political Science Review, 115:4, 2021, https://doi.org/10.1017/S0003055421000459.

59 Ryan C. Moore and Jeffrey T. Hancock. »A digital media literacy intervention for older adults improves resilience to fake news«, Scientific Reports 12:6008, 2022, https://doi.org/10.1038/s41598-022-08437-0.

60 Katherine Clayton et al. »Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media.« Political Behavior, 42:4, 2020, https://doi.org/10.1007/s11109-019-09533-0.

61 Nathaniel Sirlin, Ziv Epstein, Antonio A. Arechar, and David G. Rand. »Digital literacy is associated with more discerning accuracy judgments but not sharing intentions«, Harvard Kennedy School Misinformation Review: 2:6, 2021. https://doi.org/10.37016/mr-2020-83.

62 Jessica Gottlieb, Claire L. Adida, and Richard Moussa. »Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire«, OSF Preprints, 8 November 2022, DOI:10.31219/osf.io/6x4wy.

content creators.[63] Where grassroots efforts are concerned, activists have succeeded in challenging the attitudes and behaviour of conspiracy theorists through polite engagement and gentle questioning, although this method is time-consuming and difficult to scale upwards.[64] Regarding other direct interventions, research indicates that reminding Internet users about the limits of acceptable behaviour (for instance, not using violent language), can reduce verbal aggression.[65] There are worries that counterspeech may increase the circulation of falsehoods,[66] although inertia is also inadvisable.[67] Responding to misinformation with empathy-based counterspeech (by telling someone their language is hurtful, for example, »Your post is very painful for Jewish people to read«) appears to be more effective than other methods that use humour or warn of negative consequences.[68]

INTERVENTIONS WITH REGARD TO HATE AND RADICALISATION

One of the most direct interventions with regard to hate and the radicalisation that it can foster is that of interfaith work. This is a longstanding tradition which aims to bring people of different faiths together in order to forge connections and find common ground. Meetings of religious leaders from rival communities have, on occasion, been deemed instrumental in mitigating conflict. For example, the 2002 Alexandria Agreement has been credited with connecting the Chief Rabbi of Israel and the Mufti of Hebron, who cooperated during a political crisis to avert sectarian violence.[69] At a grassroots level, interfaith dialogue and cross-community activities, such as extending invites to lectures and family occasions, is deemed helpful in reducing tensions.[70] It is difficult, however, to determine what makes for a positive interfaith intervention, especially as it aims to change behaviour and attitudes, not just about another religion but one's

own.[71] Some of the most positive developments in interfaith or intercommunity work seem unlikely to repeat, as evidenced by the case of Mo Salah, a Muslim football player whose stint at Liverpool coincided with a substantial drop in local Islamophobic tweets compared to other regions.[72] In more quotidian examples of successful interventions, personal relationships and developing trust have been cited as the basis of effective interfaith work.[73]

A more recent approach is the teaching of values that are opposed to hate and extremism. In 2014, the UK government published guidance for schools to »strengthen the barriers to extremism« by promoting »fundamental British values«: democracy, the rule of law, individual liberty, and mutual respect and tolerance of those with different faiths and beliefs.[74] Fundamental British values are considered by some researchers to be an important element of an effective counter-extremism policy.[75] However, they have proved controversial, in part because they are ambiguous and open to interpretation,[76] but also because they are not unique to Britain.[77] Some educators have argued that the emphasis on fundamental British values has the effect of securitising education[78] and stigmatising Muslim students.[79] The policy was enacted in the wake of the Trojan Horse affair in which Birmingham school governors were sacked after a government report found they »espoused, endorsed, or failed to challenge« extremist views.[80] Some researchers have instead called for fundamental British values to be replaced by

63 Charlie Winter and Johanna Fürst. »Challenging Hate: Counter-speech Practices in Europe«, ICSR, 2018, https://icsr.info/wp-content/uploads/2018/03/ICSR-Report-Challenging-Hate-Counter-speech-Practices-in-Europe.pdf.

64 Harry Shukman. »Lewis believed Covid was a conspiracy. Now he's changed his mind.« The Post, 5 February 021, https://www.livpost.co.uk/p/lewis-believed-covid-was-a-conspiracy.

65 Michał Bilewicz et al. »Artificial intelligence against hate: Intervention reducing verbal aggression in the social network environment«, Aggressive Behavior, 47:3, 2021, https://doi.org/10.1002/ab.21948.

66 Liriam Sponholz. »Islamophobic Hate Speech: What is the Point of Counter-Speech? The Case of Oriana Fallaci and The Rage and the Pride«, Journal of Muslim Minority Affairs, 36:4, 2016, https://doi.org/10.1080/13602004.2016.1259054.

67 Sina Laubenstein and Alexander Urban. »Which types of campaign against hate and extremism on the internet work, which do not, and why?« in The OCCI Research Report, ISD Global, 2019, https://www.isdglobal.org/wp-content/uploads/2019/06/ISD-Hate-Speech-and-Radicalisation-Online-English-Draft-2.pdf.

68 Dominik Hangartner et al. »Empathy-based counterspeech can reduce racist hate speech in a social media field experiment«, PNAS, 118:50, 2021, https://doi.org/10.1073/pnas.2116310118.

69 United States Institute of Peace. »What Works? Evaluating Interfaith Dialogue Programs«, Special Report 123, July 2004, https://www.usip.org/sites/default/files/resources/sr123.pdf.

70 Ibid.

71 Woolf Institute. »How to Measure Success: A toolkit for the evaluation of interfaith engagement«, 2021, https://www.woolf.cam.ac.uk/assets/imported/Measures-of-Success-Toolkit-without-Repository.pdf.

72 Ala' Alrababa'h, William Marble, Salma Mousa, and Alexandra Siegel. »Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah On Islamophobic Behaviors and Attitudes«, Immigration Policy Lab Working Paper Series, 2019, DOI: 10.31235/osf.io/eq8ca.

73 Woolf Institute, 2021.

74 Department for Education and Lord Nash. »Guidance on promoting British values in schools published«, gov.uk/government/news/guidance-on-promoting-british-values-in-schools-published, 2014

75 Hannah Stuart. »Community Policing and Preventing Extremism: Lessons from Bradford«, Centre for the Response to Radicalisation and Terrorism Policy Paper No. 4, 2015, https://henryjacksonsociety.org/wp-content/uploads/2015/02/Community-Policing-and-Preventing-Extremism.pdf.

76 Bill Bolloten and Robin Richardson. »The Great British Values Disaster – education, security and vitriolic hate«, Institute of Race Relations, 2015, https://irr.org.uk/article/the-great-british-values-disaster-education-security-and-vitriolic-hate/.

77 Sally Tomlinson. »Fundamental British Values« in The Runnymede School Report: Race, Education and Inequality in Contemporary Britain, eds. Claire Alexander, Debbie Weekes-Bernard and Jason Arday, 2015, https://assets.website-files.com/61488f992b58e687f1108c-7c/617bccd5f0b573ea69b0b3f3_The%20School%20Report.pdf.

78 Christine Winter & China Mills. »The Psy-Security-Curriculum ensemble: British Values curriculum policy in English schools«, Journal of Education Policy, 35:1, 2020, https://doi.org/10.1080/02680939.2018.1493621.

79 Claire E. Crawford. »Promoting fundamental British values‹ in schools: a critical race perspective«, Curric Perspect, 37, 2017. DOI 10.1007/s41297-017-0029-3.

80 Peter Clarke, Education Commissioner. »Report into allegations concerning Birmingham schools arising from the ›Trojan Horse‹ letter«, July 2014, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/340526/HC_576_accessible_-.pdf.

a less exclusionary curriculum of Human Rights Education, as defined by the UN.[81] Human Rights Education promotes a universal culture of human rights that transcends borders, applies to »everyone, simply by virtue of being human«, and focuses on »equality, dignity, and common humanity«.[82]

Since 2015, staff at schools, universities, and healthcare providers have had a statutory duty to refer young people who have shown signs of violent and non-violent extremism to a designated safeguarding lead. This is known as the Prevent duty, and is a key part of the UK government's counter-terrorism strategy, which aims to stop people from being radicalised.[83] The Prevent duty has been criticised for being vague, potentially curbing free speech in classrooms, allegedly discriminating against Muslim students, and potentially complicating the safeguarding duties that teachers and healthcare workers have towards young people.[84] The perceived flaws of the Prevent duty have led to calls for it to be scrapped amid a popular assertion that the programme frequently and unfairly accuses Muslim children of extremism.[85] However, others argue that criticisms of the programme have often been made in bad faith or misreported by the media, and instead call for the programme to be clarified, arguing that it is a crucial part of Britain's counter-terrorism strategy.[86] At the time of writing, a review of Prevent has been delivered to the Home Office but has not yet been published.[87]

Finally, the »redirect method« hopes to de-radicalise Internet users who search for extreme material by providing them with alternative links that, if clicked on, will send them to content encouraging them to question their beliefs.[88] In some quarters, it has been hailed as the most effective way of putting alternative narratives in front of an online audience, some of whose members may be radicalised.[89] Assessing the impact of the redirect method is difficult, as interest in extreme material does not necessarily denote desire to join an extremist group, nor does clicking on the redirected content equate to a change in attitude or behaviour.[90] A 2018 report found that there had been no »assess[ment of] the impact of [redirect method] video content on user attitudes or behaviour«[91], and noted that click rates ranged from 2.22–3.19 % and led to average watch times of just 10–59 seconds from the small minority of targeted web users who actually did click through: hardly enough time, one might suppose, to lead extremists towards meaningful questioning of their worldviews. The redirect method also relies entirely on the quality of the content towards which Internet users are sent, which, if poorly curated, might exacerbate any problematic views which they might hold. For example, a leading tech startup specialising in the redirect method was in one case found to have redirected Internet users to the YouTube channel of »a convicted felon and disinformation« amplifier who espouses violent antisemitic, anti-cop, and anarchist views«.[92]

CONCLUSION AND SCOPE FOR FURTHER WORK

It is clear that radicalisation, hate speech, and misinformation now reach substantial audiences online. However, just as there is evidence that offline contact with other extremists is usually necessary in order to transform a mere keyboard warrior into a dangerous terrorist, there is evidence that meaningful offline interaction may be necessary to draw people away from extremism. That is, while there is no one method that is completely effective, interventions that involve real-world human contact seem to be most impactful, as do ones that plan for long-term, repeated engagement. The long tradition of interfaith dialogue may provide a model here, provided that it can be adapted to bring together more communities than those defined simply by faith. An online dimension to such work will certainly be necessary, because people today – and especially the young – live in a social world that is substantially mediated by digital technology. But purely technological fixes, such as the uncertain »redirect method«, should probably be viewed with scepticism.

81   United Nations Declaration on Human Rights Education and Training, 2011, https://www.ohchr.org/en/resources/educators/human-rights-education-training/11-united-nations-declaration-human-rights-education-and-training-2011.

82   Diane Webber and Alison Struthers. »Critiquing Approaches to Countering Extremism: The Fundamental British Values Problem in English Formal Schooling«, Commission for Countering Extremism, 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/834353/Webber-Struthers-Critiquing-Approaches-to-Countering-Extremism.pdf.

83   Home Office. »Revised Prevent duty guidance: for England and Wales«, 1 April 2021, https://www.gov.uk/government/publications/prevent-duty-guidance/revised-prevent-duty-guidance-for-england-and-wales.

84   Anne-Lynn Dudenhoefer. »Resisting Radicalisation: A Critical Analysis of the UK Prevent Duty«, Journal for Deradicalization, 14, Spring 2018, https://journals.sfu.ca/jd/index.php/jd/article/view/138.

85   Aislinn O'Donnell. »Securitisation, Counterterrorism and the Silencing of Dissent: The Educational Implications of Prevent«, British Journal of Education Studies, 64:1, 2016, http://dx.doi.org/10.1080/00071005.2015.1121201.

86   Hannah Stuart. »Counter-Terrorism and Extremism in Great Britain Since 7/7«, Counter Extremism Group, 1, July 2020, https://counterextremism.org.uk/wp-content/uploads/2020/07/0001_CT_CE_since_7_7_05_WEB_1_6.pdf.

87   Laura Hutchinson. »The Problem With Prevent«, The House Magazine, 22 November 2022, https://www.politicshome.com/thehouse/article/the-problem-with-prevent.

88   Christina Schori Liang and Matthew John Cross. »White Crusade: How to Prevent Right-Wing Extremists from Exploiting the Internet«, Geneva Centre for Security Policy, 11, July 2020, https://dam.gcsp.ch/files/doc/white-crusade-how-to-prevent-right-wing-extremists-from-exploiting-the-internet.

89   Alexander Meleagrou-Hitchens. »The Challenges and Limitations of Online Counter-Narratives in the Fight against ISIS Recruitment in Europe and North America«, Georgetown Journal of International Affairs, 18:3, 2017.

90   Meleagrou-Hitchens, 2017; Erin Saltman, Farshad Kooti and Karly Vockery. »New Models for Deploying Counterspeech: Measuring Behavioral Change and Sentiment Analysis«, Studies in Conflict & Terrorism, DOI: 10.1080/1057610X.2021.1888404.

91   Todd C. Helmus and Kurt Klein. »Assessing Outcomes of Online Campaigns Countering Violent Extremism: A Case Study of the Redirect Method«, RAND Corporation, 2018, https://doi.org/10.7249/RR2813.

92   »A Case Study in Redirection«, Network Contagion Research Institute, 2021, https://networkcontagion.us/wp-content/uploads/a-case-study-in-redirection.pdf.

What kinds of interventions can be envisaged? An under-explored area, especially given the »top-down« nature of much counter-misinformation, counter-extremism, and counter-hate work, is that of co-creation of materials or learning experiences with members of specific communities, for distribution within those communities. Such an approach would allow for repeated re-creation and updating of materials as the cultural landscape changes, and – if properly supervised – could avoid the problem of low quality or poorly-curated content. Moreover, the process of co-creation could itself form part of an inter-group dialogue, leading to mutual exploration of misinformation beliefs, and building capacity for critical reasoning and meta-cognition by providing participants with an opportunity to arrest the formation of snap judgements about social out-groups and subsequently to reflect on the basis for such judgement in the presence of members of those groups who are likely to exhibit multiple points of personal and ideological similarity to present in-group members, as well as multiple points of personal and ideological difference from the supposedly representative figures familiar from social media. This latter point is likely to prove vitally important for the fostering of empathy in the face of the extreme polarisation sadly apparent in online discussions, producing the »counterbalancing effects of local heterogeneity« whose absence in non-local digital interaction a recent study argues to explain the outraged and unreasonable tone of much online debate.[93] A case can be made that community, educational, religious, and cultural organisations would be better suited to facilitating such a project than the state agencies, tech companies, and NGOs that currently dominate the counter-misinformation, counter-hate, and counter-extremism space.

---

93 Petter Törnberg, »How Digital Media Drive Affective Polarization through Partisan Sorting«, *PNAS* 119, no. 42 (2022): 1, https://doi.org/10.1073/pnas.2207159119.

REFERENCES

Allington, Daniel, and Tanvi Joshi. '"What Others Dare Not Say": An Antisemitic Conspiracy Fantasy and Its YouTube Audience'. *Journal of Contemporary Antisemitism* 3, no. 1 (2020): 35–53. https://doi.org/10.26613/jca/3.1.42.

Allington, Daniel, Siobhan McAndrew, Bobby Duffy, and Vivienne Louisa Moxham-Hall. 'Trust and Experiences of National Health Service Healthcare Do Not Fully Explain Demographic Disparities in Coronavirus Vaccination Uptake in the UK: A Cross--sectional Study'. *BMJ Open* 12, no. e053827 (2022). https://doi.org/10.1136/ bmjopen-2021-053827.

Allington, Daniel, Siobhan McAndrew, Vivienne Louisa Moxham-Hall, and Bobby Duffy. 'Coronavirus Conspiracy Suspicions, General Vaccine Attitudes, Trust, and Coronavirus Information Source as Predictors of Vaccine Hesitancy among UK Residents during the COVID-19 Pandemic'. *Psychological Medicine*, 2021. https://doi.org/10.1017/S0033291721001434.

——. 'Media Usage Predicts Intention to Be Vaccinated against SARS-CoV-2 in the US and the UK'. *Vaccine* 39 (2021): 2595–2603. https://doi.org/10.1016/j.vaccine.2021.02.054.

Alqaisya, Walaa. 'Palestine and the Will to Theorise Decolonial Queering'. *Middle East Critique* 29, no. 1 (2020): 87–113. https://doi.org/10.1080/19436149.2020.1704505.

Alrababa'h, Ala', William Marble, Salma Mousa, and Alexandra Siegel. 'Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes'. *American Political Science Review* 115, no. 4 (2021): 1111–28. https://doi.org/10.31235/osf.io/eq8ca.

Askanius, Tina. 'On Frogs, Monkeys, and Execution Memes: Exploring the Humor-Hate Nexus at the Intersection of Neo-Nazi and Alt-Right Movements in Sweden'. *Television and New Media* 22, no. 2 (2021): 147–65. https://doi.org/10.1177/1527476420982234.

Badrinathan, Sumitra. 'Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India'. *American Political Science Review* 115, no. 4 (2021): 1325–41. https://doi.org/10.1017/S0003055421000459.

Banas, John A., and Stephen A. Rains. 'A Meta-Analysis of Research on Inoculation Theory'. *Communication Monographs* 77, no. 3 (2010): 281–311. https://doi.org/10.1080/03637751003758193.

Barta, Tony. 'Discourses of Genocide in Germany and Australia: A Linked History'. *Aboriginal History* 25 (2001): 37–56.

Basol, M., Jon Roozenbeek, and S. van der Linden. 'Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News'. *Journal of Cognition* 3, no. 1 (2020). https://doi.org/10.5334/joc.91.

BBC. 'Pro-Palestine Rally: Racial Hatred Charges Dropped for Two More Men'. *BBC Local News*, 21 November 2022. https://www.bbc.co.uk/news/uk-england-london-63707451.

Bilewicz, Michał, Patrycja Tempska, Leliwa Gniewosz, Maria Dowgiałło, Michalina Tańska, Rafał Urbaniak, and Michał Wroczyński. 'Artificial Intelligence against Hate: Intervention Reducing Verbal Aggression in the Social Network Environment'. *Aggressive Behaviour* 47, no. 3 (2021): 260–66.

Bjørgo, Tore. 'Militant Neo-Nazism in Sweden'. *Terrorism and Political Violence* 5, no. 3 (1993): 1556–1836. https://doi.org/10.1080/09546559308427219.

Bolloten, Bill, and Robin Richardson. 'The Great British Values Disaster – Education, Security and Vitriolic Hate'. *Institute of Race Relations* (blog), 2015. https://irr.org.uk/article/the-great-british-values-disaster-education-security-and-vitriolic-hate/.

Bolton, Matt. 'Conceptual Vandalism, Historical Distortion: The Labour Antisemitism Crisis and the Limits of Class Instrumentalism'. *Journal of Contemporary Antisemitism* 3, no. 2 (2020): 11–30. https://doi.org/10.26613/jca/3.2.56.

Callaghan, Timothy, Ali Moghtaderi, Jennifer A. Lueck, Peter Hotez, Ulrich Strych, Avi Dor, Erika Franklin Fowler, and Matthew Motta. 'Correlates and Disparities of COVID-19 Vaccine Hesitancy'. *Social Science & Medicine* 272 (2021): 113638.

Cascini, Fidelia, Ana Pantovic, Yazan Al-Ajlouni, Giovanna Failla, Valeria Puleo, Andriy Melnyk, Alberto Lontano, and Walter Ricciardi. 'Social Media and Attitudes towards a COVID-19 Vaccination: A Systematic Review of the Literature'. *EClinicalMedicine* 48 (2022): 101454. https://doi.org/10.1016/j.eclinm.2022.101454.

CCE. 'Challenging Hateful Extremism'. London: Commission for Countering Extremism, 7 October 2019. https://www.gov.uk/government/publications/challenging-hateful-extremism.

Christensen, Dino P., Sarah E. Kreps, and Douglas L. Kriner. 'Contemporary Presidency: Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion'. *Presidential Studies Quarterly* 51, no. 1 (2020): 151–65. https://doi.org/10.1111/psq.12687.

Clarke, Peter. 'Report into Allegations Concerning Birmingham Schools Arising from the "Trojan Horse" Letter'. London: Department for Education, 2014. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/340526/HC_576_accessible_-.pdf.

Clayton, Katherin, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, et al. 'Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media'. *Political Behavior* 42 (2020): 1073–95. https://doi.org/10.1007/s11109-019-09533-0.

Cook, John, Stephan Lewandowsky, and Ullrich K.H. Ecker. 'Neutralizing Misinformation through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence'. *PLoS ONE* 12, no. 5 (2017). https://doi.org/10.1371/journal.pone.0175799.

Crawford, Claire E. 'Promoting "Fundamental British Values" in Schools: A Critical Race Perspective'. *Curriculum Perspectives* 37 (2017): 197–204. https://doi.org/10.1007/s41297-017-0029-3.

Department for Education, and Lord Nash. 'Guidance on Promoting British Values in Schools Published: Strengthened Guidance on Improving the Spiritual, Moral, Social and Cultural Development of Pupils.' *UK Government News* (blog), 27 November 2014. https://www.gov.uk/government/news/guidance-on-promoting-british-values-in-schools-published.

Dudenhoefer, Anne-Lynne. 'Resisting Radicalisation: A Critical Analysis of the UK Prevent Duty'. *Journal for Deradicalization* 14 (2018): 153–91.

Edthofer, Julia. 'Israel as Neo-Colonial Signifier? Challenging de-Colonial Anti-Zionism'. *Journal for the Study of Antisemitism* 7 (2015): 31–51.

Ekman, Mattias. 'The Great Replacement: Strategic Mainstreaming of Far-Right Conspiracy Claims'. *Convergence: The International Journal of Research into New Media Technologies* 28, no. 4 (2022): 1127–43. https://doi.org/10.1177/13548565221091983.

European Commission. 'Report on the Implementation of the Action Plan against Disinformation'. Brussels: Publications Office of the European Union, 2019. https://op.europa.eu/en/publication-detail/-/publication/8a94fd8f-8e92-11e9-9369-01aa75ed71a1/language-en.

General Assembly of the United Nations. United Nations Declaration on Human Rights Education and Training (2011), Resolution 66/137, A/RES/66/137 § (2011). https://www.ohchr.org/en/resources/educators/human-rights-education-training/11-united-nations-declaration-human-rights-education-and-training-2011.

Gottlieb, Jessica, Claire L. Adida, and Richard Moussa. 'Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire'. OSF Preprints, 8 November 2022. https://doi.org/10.31219/osf.io/6x4wy.

Grierson, Jamie. 'Anti-Asian Hate Crimes up 21% in UK during Coronavirus Crisis: Ministers Also Tell Commons Committee Revenge Porn and Online Fraud Have Increased'. *The Guardian*. 13 May 2020. https://www.theguardian.com/world/2020/may/13/anti-asian-hate-crimes-up-21-in-uk-during-coronavirus-crisis.

Guess, Andrew M., Michael Lerner, and Benjamin Lyons. 'A Digital Media Literacy Intervention Increases Discernment between Mainstream and False News in the United States and India'. *PNAS* 117, no. 27 (2019): 15536–45. https://doi.org/10.1073/pnas.1920498117.

Hall, Jonathan. 'Keyboard Warriors or International Terrorists?' London: Independent Reviewer of Terrorism Legislation, 14 July 2022. https://terrorismlegislationreviewer.independent.gov.uk/wp-content/uploads/2022/07/IRTL-Keyboard.pdf.

Hangartner, Dominik, Gloria Gennaro, Sary Alasiri, Nicholas Bahrich, Alexandra Bornhoft, Joseph Boucher, Buket Buse Demirci, et al. 'Empathy-Based Counterspeech Can Reduce Racist Hate Speech in a Social Media Field Experiment'. *PNAS* 118, no. 50 (2021): e2116310118. https://doi.org/10.1073/pnas.2116310118.

Helmus, Todd C., and Kurt Klein. 'Assessing Outcomes of Online Campaigns Countering Violent Extremism: A Case Study of the Redirect Method'. RAND Corporation, 2018. https://doi.org/10.7249/RR2813.

Hern, Alex. '5G Conspiracy Theories Fuel Attacks on Telecoms Workers: One Openreach Engineer Reports Being Forced to Flee Angry Group since Covid-19 Crisis'. *The Guardian*. 7 May 2020. https://www.theguardian.com/business/2020/may/07/5g-conspiracy-theories-attacks-telecoms-covid.

Hirsh, David. 'How the Word "Zionist" Functions in Antisemitic Vocabulary'. *Journal of Contemporary Antisemitism* 4, no. 2 (2021): 1–17.

Home Office. 'Revised Prevent Duty Guidance: For England and Wales'. UK Government, 2021. https://www.gov.uk/government/publications/prevent-duty-guidance/revised-prevent-duty-guidance-for-england-and-wales.

Iyengar, Ananya, Poorvi Gupta, and Nidhi Priya. 'Inoculation against Conspiracy Theories: A Consumer Side Approach to India's Fake News Problem'. *Applied Cognitive Psychology Early View* (2022). https://doi.org/10.1002/acp.3995.

Jaspal, Rusi. *Antisemitism and Anti-Zionism: Representation, Cognition, and Everyday Talk*. Abingdon: Routledge, 2014.

Jolley, Daniel, and Karen M Douglas. 'Prevention Is Better than Cure: Addressing Anti-Vaccine Conspiracy Theories'. *Journal of Applied Social Psychology* 47 (2017): 459–69. https://doi.org/10.1111/jasp.12453.

Jupskås, Anders Ravik, Ravndal, Jacob Aasland, Charlotte Tandberg, and Madeleine Thorstensen. 'Five Myths of Far-Right Violence: Improved Systematic Data on Severe Far-Right Violence in Western Europe Debunks Five Common Myths in the Field'. *Center for Research on Extremism* (blog), 23 November 2022. https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2022/five-myths-of-far-right-violence.html.

Kenyon, Jonathan, Jens Binder, and Christopher Baker-Beall. 'Exploring the Role of the Internet in Radicalisation and Offending of Convicted Extremists'. London: Ministry of Justice, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1017413/exploring-role-internet-radicalisation.pdf.

—. 'The Internet and Radicalisation Pathways: Technological Advances, Relevance of Mental Health and Role of Attackers'. Ministry of Justice, 2022. https://www.gov.uk/government/publications/internet-and-radicalisation-pathways-technological-advances-relevance-of-mental-health-and-role-of-attackers.

Kim, Hyunuk, and Dylan Walker. 'Leveraging Volunteer Fact Checking to Identify Misinformation about COVID-19 in Social Media'. *Harvard Kennedy School Misinformation Review*, 2020. https://doi.org/doi.org/10.37016/mr-2020-021.

Laubenstein, Sina, and Alexander Urban. 'Which Types of Campaign against Hate and Extremism on the Internet Work, Which Do Not, and Why?' In *Hate Speech and Radicalisation Online*: *The OCCI Research Report*, 50–57. London / Washington DC / Beirut / Toronto: Institute for Strategic Dialogue, 2019.

Liang, Christina Schori, and Matthew John Cross. 'White Crusade: How to Prevent Right-Wing Extremists from Exploiting the Internet'. Geneva: Geneva Centre for Security Policy, 2020. https://dam.gcsp.ch/files/doc/white-crusade-how-to-prevent-right-wing-extremists-from-exploiting-the-internet.

Margolin, Drew B., Aniko Hannak, and Ingmar Weber. 'Political Fact-Checking on Twitter: When Do Corrections Have an Effect?' *Political Communication* 35, no. 2 (2018): 196–219. https://doi.org/10.1080/10584609.2017.1334018.

Meleagrou-Hitchens, Alexander. 'The Challenges and Limitations of Online Counter-Narratives in the Fight against ISIS Recruitment in Europe and North America'. *Georgetown Journal of International Affairs* 18, no. 3 (2017): 95–104.

Moore, Ryan C., and Jeffrey T. Hancock. 'A Digital Media Literacy Intervention for Older Adults Improves Resilience to Fake News'. Scientific Reports 12 (2022): 6008. https://doi.org/10.1038/s41598-022-08437-0.

Moss, Molly. 'Coronation Street Confirms Max Turner to Be Groomed by Extremists: The Storyline Will Examine How Extremist Groups Target, Groom and Operate'. *Radio Times*, 27 October 2022. https://www.radiotimes.com/tv/soaps/coronation-street/coronation-street-max-turner-grooming-storyline-newsupdate/.

Murphy, Jamie, Frédérique Vallières, Richard P Bentall, Mark Shevlin, Orla McBride, Todd Hartman, Ryan McKay, et al. 'Psychological Characteristics Associated with COVID-19 Vaccine Hesitancy and Resistance in Ireland and the United Kingdom'. *Nature Communications* 12, no. 29 (5 June 2020). https://doi.org/10.1038/s41467-020-20226-9.

Nassetta, Jack, and Kimberly Gross. 'State Media Warning Labels Can Counteract the Effects of Foreign Misinformation'. *Harvard Kennedy School Misinformation Review*, 2020. https://doi.org/10.37016/mr-2020-45.

NCRI. 'A Case Study in Redirection'. Network Contagion Research Institute, 2021. https://networkcontagion.us/wp-content/uploads/a-case-study-in-redirection.pdf.

Nygren, Thomas, Mona Guath, Carl-Anton Werner Axelsson, and Divina Frau-Meigs. 'Combatting Visual Fake News with a Professional Fact-Checking Tool in Education in France, Romania, Spain and Sweden'. *Information* 12, no. 5 (2021). https://doi.org/10.3390/info12050201.

Obaidi, Milan, Jonas Kunst, Simon Ozer, and Sasha Y. Kimel. 'The "Great Replacement" Conspiracy: How the Perceived Ousting of Whites Can Evoke Violent Extremism and Islamophobia'. *Group Processes and Intergroup Relations* 25, no. 7 (2022): 1675–95. https://doi.org/10.1177/13684302211028293.

O'Donnell, Aislinn. 'Securitisation, Counterterrorism and the Silencing of Dissent: The Educational Implications of Prevent'. *British Journal of Educational Studies* 64, no. 1 (2016): 53–76. https://doi.org/10.1080/00071005.2015.1121201.

Patterns of Prejudice. '"Anti-Zionists" and Antisemites'. *Patterns of Prejudice* 4, no. 4 (1970): 28.

Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. 'The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings'. *Management Science* 66, no. 11 (2020): 4944–57. https://doi.org/10.1287/mnsc.2019.3478.

Pennycook, Gordon, Tyrone D. Cannon, and David G. Rand. 'Prior Exposure Increases Perceived Accuracy of Fake News'. *Journal of Experimental Psychology* 147, no. 12 (2018): 1865–80. https://doi.org/10.1037/xge0000465.

Potter, Ethan, and Thomas J. Wood. 'The Global Effectiveness of Fact-Checking: Evidence from Simultaneous Experiments in Argentina, Nigeria, South Africa, and the United Kingdom'. *PNAS* 118, no. 37 (2021). https://doi.org/10.1073/pnas.2104235118.

Roozenbeek, Jon, and Sander van der Linden. 'Fake News Game Confers Psychological Resistance against Online Misinformation'. *Palgrave Communications* 5 (2019): 65. https://doi.org/10.1057/s41599-019-0279-9.

Roozenbeek, Jon, Sander van der Linden, and Thomas Nygren. 'Prebunking Interventions Based on "Inoculation" Theory Can Reduce Susceptibility to Misinformation across Cultures'. *Harvard Kennedy School Misinformation Review* 1, no. 2 (2020): 1–23. https://doi.org/10.37016//mr-2020-008.

Sabbagh, Dan. 'MI5 Chief: UK Will Have to Tackle Russian Aggression "for Years to Come"'. *Guardian*, 16 November 2022. https://www.the-guardian.com/uk-news/2022/nov/16/mi5-chief-uk-will-have-to-tackle-russian-aggression-for-years-to-come.

Saltman, Erin, Farshad Kooti, and Karly Vockery. 'New Models for Deploying Counterspeech: Measuring Behavioral Change and Sentiment Analysis'. Studies in Conflict & Terrorism, 2020. https://doi.org/10.1080/1057610X.2021.1888404.

Sharevski, Filipo, Raniem Alsaadi, Peter Jachim, and Emma Pieroni. 'Misinformation Warnings: Twitter's Soft Moderation Effects on COVID-19 Vaccine Belief Echoes'. *Computer Security* 114, no. 102577 (2022). https://doi.org/10.1016/j.cose.2021.102577.

Shukman, Harry. 'Lewis Believed Covid Was a Conspiracy. Now He's Changed His Mind'. *The Post*. 5 February 2022. https://www.livpost.co.uk/p/lewis-believed-covid-was-a-conspiracy.

Sirlin, Nathaniel, Ziv Epstein, Antonio Arechar, and David G. Rand. 'Digital Literacy Is Associated with More Discerning Accuracy Judgments but Not Sharing Intentions'. *Harvard Kennedy School Misinformation Review* 2, no. 6 (2021). https://doi.org/10.37016/mr-2020-83.

Skafle, Ingjerd, Anders Nordahl-Hansen, Daniel S. Quintana, Rolf Wynn, and Elia Gabarron. 'Misinformation about COVID-19 Vaccines on Social Media: Rapid Review'. *Journal of Medical Internet Research* 24, no. 8 (2022): e37367. https://doi.org/10.2196/37367.

Sponholz, Liriam. 'Islamophobic Hate Speech: What Is the Point of Counter-Speech? The Case of Oriana Fallaci and The Rage and the Pride'. *Journal of Muslim Minority Affairs* 36, no. 4 (2016): 502–22. https://doi.org/10.1080/13602004.2016.1259054.

Stuart, Hannah. 'Community Policing and Preventing Extremism: Lessons from Bradford'. London: Centre for the Response to Radicalisation and Terrorism, 2015. https://henryjacksonsociety.org/wp-content/uploads/2015/02/Community-Policing-and-Preventing-Extremism.pdf.

—. 'Counter-Terrorism and Extremism in Great Britain since 7/7'. London: Counter Extremism Group, 2020. https://counterextremism.org.uk/wp-content/uploads/2020/07/0001_CT_CE_since_7_7_05_WEB_1_6.pdf.

Swerling, Gabriella. 'Men Arrested over Pro-Palestinian Convoy Linked to Previous Anti-Semitic Incident: It Is Alleged That Two Vehicles Followed a Jewish Man in Salford before the Passengers Exited Their Cars and Attempted to Smash His Windows'. *Daily Telegraph*. 18 May 2021. https://www.telegraph.co.uk/news/2021/05/18/men-arrested-pro-palestine-london-convoy-linked-previous-anti/.

Terrorism Act (2000). https://www.legislation.gov.uk/ukpga/2000/11/section/58.

Thorson, Emily. 'Belief Echoes: The Persistent Effects of Corrected Misinformation'. *Political Communication* 33, no. 3 (2016). https://doi.org/10.1080/10584609.2015.1102187.

Tomlinson, Sally. In *The Runnymede School Report: Race, Education, and Inequality in Contemporary Britain*, edited by Claire Alexander, Debbie Weekes-Bernard, and Jason Arday, 10–13. London: Runnymede, 2015. https://assets.website-files.com/61488f992b58e687f1108c-7c/617bccd5f0b573ea69b0b3f3_The%20School%20Report.pdf.

Törnberg, Petter. 'How Digital Media Drive Affective Polarization through Partisan Sorting'. *PNAS* 119, no. 42 (2022): e2207159119. https://doi.org/10.1073/pnas.2207159119.

UK Government. 'Disinformation and "Fake News": Interim Report: Government Response to the Committee's Fifth Report'. London: UK Government, 2018. https://publications.parliament.uk/pa/cm201719/cm-select/cmcumeds/1630/163002.htm.

USIP. 'What Works? Evaluating Interfaith Dialogue Programs'. Washington D.C.: United States Institute for Peace, 2004. https://www.usip.org/sites/default/files/resources/sr123.pdf.

Waterson, Jim, and Alex Hern. 'At Least 20 UK Phone Masts Vandalised over False 5G Coronavirus Claims: Industry Body Assures People in Open Letter There Is No Link between 5G and Pandemic'. *The Guardian*. 6 April 2020. https://www.theguardian.com/technology/2020/apr/06/at-least-20-uk-phone-masts-vandalised-over-false-5g-coronavirus-claims.

Webber, Diana, and Alison Struthers. 'Critiquing Approaches to Countering Extremism: The Fundamental British Values Problem in English Formal Schooling'. London: Commission for Countering Extremism, 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/834353/Webber-Struthers-Critiquing-Approaches-to-Countering-Extremism.pdf.

Winter, Charlie, and Johanna Fürst. 'Challenging Hate: Counter-Speech Practices in Europe'. London: International Centre for the Study of Radicalisation and Political Violence, King's College London, 2017. https://icsr.info/wp-content/uploads/2018/03/ICSR-Report-Challenging-Hate-Counter-speech-Practices-in-Europe.pdf.

Woolf Institute. 'How to Measure Success: A Toolkit for the Evaluation of Interfaith Engagement'. Cambridge: University of Cambridge, 2021. https://www.woolf.cam.ac.uk/assets/imported/Measures-of-Success-Toolkit-without-Repository.pdf.

Zhang, Jingwen, Jieyu Ding Featherstone, Christopher Calabrese, and Magdalena Wojcieszak. 'Effects of Fact-Checking Social Media Vaccine Misinformation on Attitudes toward Vaccines'. *Preventative Medicine* 145, no. article 106408 (2021). https://doi.org/10.1016/j.ypmed.2020.106408.

ABOUT THE AUTHORS

Dr Daniel Allington is Reader in Social Analytics at King's College London. He has published extensively on conspiracy theories and hate speech, and has worked both with the UK Commission for Countering Extremism and with the UK government's independent advisers on antisemitism and political violence.

Harry Shukman is a freelance journalist and writes *Scout*, a blog about conspiracy theories and far-right extremism. He is a former news reporter at *The Times* and recently finished a master's degree in Terrorism, Security and Society at King's College London.

Rachel Briscoe is a writer, dramaturg and facilitator. She is co-founder lead artist of *Fast Familiar*, a digital story studio who make audience-centric experiences which are playful, participatory and political.

William Galinsky is an artistic director, theatre-maker and producer. He works as Director of Programming for JW3 and has a special interest in using cultural participation to address shared civil society issues.

© 2023

FES LONDON OFFICE

The London Office is part of the international network of FES. It was established in 1988 to promote better understanding of British – German relations, mainly by means of seminars and reports on political trends in the UK and Germany.

The Friedrich-Ebert-Stiftung's work in the United Kingdom focuses in particular on the exchange of ideas and discussion on the following topics:

– common challenges facing Germany, the United Kingdom and the European Union

– economic and social policy
– experiences gained from the differing regional and local policies and practices
– a continuing dialogue between politicians in both countries, particularly through a series of meetings for MPs
– a continuing dialogue between the trades unions, particularly via the annual British-German Trade Union Forum launched in 2002

For more information please visit:
www.uk.fes.de

www.fes.de/bibliothek/fes-publikationen

CANARIES IN THE COALMINE

Building Resilience to Online Misinformation among Young People in the UK

→

Online misinformation reaches substantial youth audiences, prompting concerns that it can exacerbate inter-group hate and radicalisation. Belief in misinformation, such as the Great Replacement conspiracy theory, has been cited as a possible motive for mass shootings in the US and Europe. Vaccine hesitancy and the vandalisation of 5G masts has additionally been linked with acceptance of misinformation regarding COVID-19.

→

Tackling this behaviour has become a source of great worry, and governments, religious organisations and academic researchers suggest a number of interventions. Efforts to curb misinformation that exist solely online — such as fact-checking services and content warnings — likely have a limited impact and cost a great deal. The goal of achieving a durable impact is also challenging: positive effects of misinformation interventions seem to wane with time.

→

No single method is completely effective, although we argue that the most impactful interventions are likely to be in-person, real-world efforts drawing on the tradition of interfaith dialogue. Forging connections, finding common ground and forcing participants to examine attitudes towards other communities — as well as their own — has been linked to tangible positive outcomes. The authors suggest that socially engaged arts and cultural projects between communities may provide fruitful new avenues of exploration in this area.

Further information on the topic can be found here:
https://uk.fes.de/

FRIEDRICH EBERT STIFTUNG