CSR

Misinformation

CONTENTS

FROM THE EDITORS

Misinformation isn't a new development, but its historical presence should not downplay its impact today. Despite unprecedented access to information, the surge of misinformation, with its far-reaching implications, has become a formidable and pressing challenge that transcends borders, cultures, and languages.

This bumper issue of *CREST Security Review* highlights the role of behavioural and social sciences in battling against the relentless spread of disinformation. In an era where information is both a weapon and shield, Innes *et al.* (p. 4) starts us off with the intricate dance between open-source intelligence (OSINT) and disinformation.

Venturing into the digital, Buchanan et al. (p. 14) delves into complex motives of why people share false political information online, Dance (p. 16) confronts the pivotal role of algorithms in perpetuating disinformation and Mckenzie investigates the difficulties in discerning fake social media profiles (p. 20). On the topic of misleading, Reeve looks at the disruptive influence of false information and distraction on group dynamics (p. 12).

In the shadowy corners of misinformation, conspiracy theories lurk. Green *et al.* (p. 24) dissects the ways in which these narratives can be a catalyst for political violence, while Rottweiler & Gill (p. 8) focuses on the amplifying effect of such beliefs on societal and group-level risk factors. Orpen's doctoral research may help as she sets out her innovative 'Fuzzy' methodological approach on page 18.

In our quest for truth, building resilience to misinformation becomes paramount. Pavlounis & Davis discuss how to cultivate digital media literacy (p. 22), Ecker *et al.* explore psychological strategies that can be wielded in the battle to debunk false information, while van der Linden on page 26 draws from his book: 'Foolproof: Why We Fall for Misinformation and How to Build Immunity'. Our topic concludes with an A-Z of misinformation, illustrating how false or misleading information can be spread and ways to combat it (p. 36).

Beyond the topic of misinformation Power *et al.* (p. 34) explains the critical role of psychology in fostering teamwork among emergency services and in the realm of Countering Violent Extremism (CVE), Marsden & Lewis dig into the lesser-explored territory of how case management tools function (p. 30).

You can find the research that underpins all our articles and further reading in the 'Read More' section on page 40. Please let us know what you liked (or didn't) about this issue and what you would like to see featured in future issues — your feedback is important to us! Fill in the survey via the link or QR code on this page. Thank you.

**Rebecca Stevens & Kayleigh Brennan
Editors, *CSR.***

HELEN INNES, ANDREW DAWSON & MARTIN INNES

OSINT VS DISINFORMATION: THE INFORMATION THREATS 'ARMS RACE'

Exploring the interplay between open-source intelligence (OSINT) and disinformation to illuminate how they drive vital innovations in the organisation and conduct of each other.

Disinformation has emerged as a compelling policy problem over the past decade. Since the discovery that the St. Petersburg based Internet Research Agency attempted to interfere in the 2016 US Presidential election, multiple studies have documented various disinforming, distorting and deceptive communications shaping public understanding and political decision-making across policy domains. These include democratic elections, public health crises, climate change, counterterrorism, and warfare, amongst others. The public 'unmasking' of disinformation often relies upon a range of methods and techniques collectively labelled as 'OSINT', or open-source intelligence.
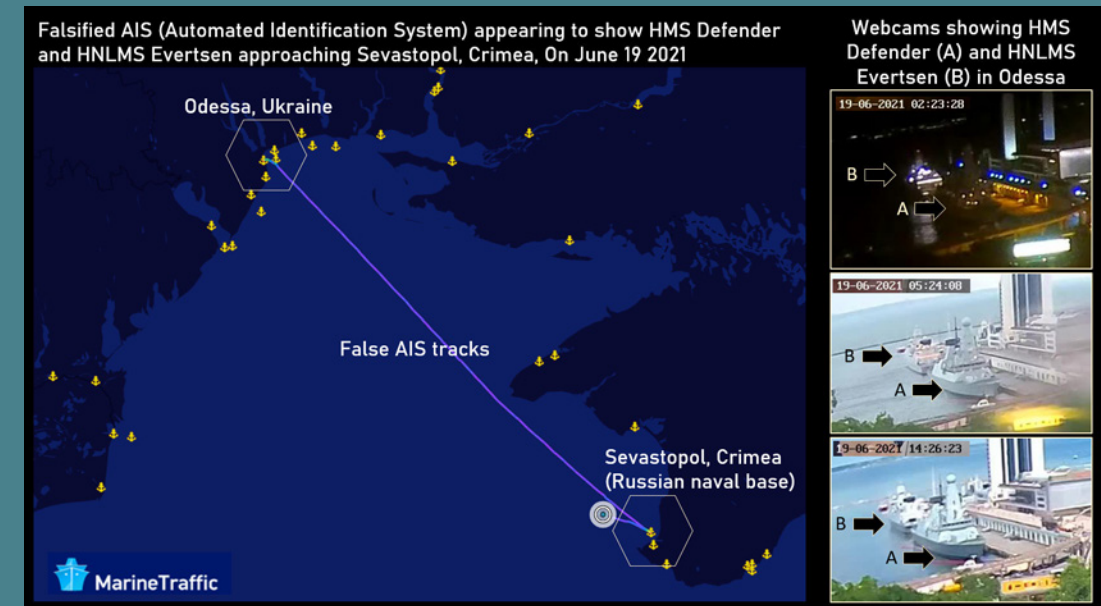
Disinformation involves communications deliberately designed and delivered to mislead. It is closely aligned with several overlapping concepts, including 'misinformation' (unintentionally misleading messages), propaganda, and conspiracy theories. Concerns about the causes and consequences of misleading public communications are not new – the term misinformation was used in the 17th century in the context of the English Civil War, and George Orwell addressed its influence when writing about the Spanish Civil War in the 1930s. The key difference today is that our information environment enables misleading, yet highly persuasive, communications to be transmitted and received at a previously unimaginable pace and scale. As a result, disinformation is an important component of hostile state information operations and comparable influence campaigns by non-state actors.

Since disinformation and OSINT are both prominent features of the contemporary information environment, it is surprising that more attention has not focused on their interplay. Instead, many empirical accounts (of varying quality and sophistication) now describe various information operations and disinformation campaigns. However, these are largely separate from an increasing number of books on the craft of open-source intelligence collection and analysis.

> **The public 'unmasking' of disinformation often relies upon a range of methods and techniques collectively labelled as 'OSINT', or open-source intelligence.**

There are intriguing co-production processes regarding how disinformation and OSINT shape innovations in each other. There is a kind of 'arms race' between OSINT analysts and the authors of disinformation. The purveyors of disinforming, distorting and deceptive communications seek to construct messages that reach and impact their targets, obscuring their origins and circumventing any attempts to intercept them. Meanwhile, the open-source analyst community seek to configure methodologies that maximise the chances of discovering misleading messages and confidently attributing sources.

There is then a continual dance of point and counterpoint as each side seeks to outwit and out-flank the other. The result is that disinformation frequently evolves and adapts, seeking new opportunities for malign influence whilst dodging the defences erected against it. The significance of this is twofold. First, although public and political discourse around disinformation centres on the role of social media, there are other vectors via which it can be transmitted and received. Second, as implied above, crises like the war in Ukraine can act as a crucible of innovation, inducing quick and substantial breakthroughs in deceptive communications. We briefly explore two examples of these dynamics.

Falsified AIS (Automated Identification System) appearing to show HMS Defender and HNLMS Evertsen approaching Sevastopol, Crimea, On June 19 2021. Webcams showing HMS Defender (A) and HNLMS Evertsen (B) in Odessa.

AIS SPOOFING

Automatic Identification System (AIS) is a radio-based system designed to alert ships to other ships in their area, preventing collisions due to poor visibility. An AIS transponder receives data from a GPS to broadcast the ship's position whilst receiving similar messages from AIS transponders on other ships in the area, allowing all the ships and their headings to be plotted on a map. Open-source marine traffic aggregators such as 'MarineTraffic.com' and 'VesselFinder.com' use AIS transponder messages to create global real-time maps of ship movements. To do this, they rely on volunteers erecting antennas on the coastline to receive AIS signals from passing ships, which are decoded by a computer and uploaded to the website. AIS is not encrypted and was not designed with security in mind. As such, AIS signals can be spoofed, resulting in incorrect or missing AIS data.

On 19 June 2021, two NATO warships were recorded on MarineTraffic.com leaving Odesa in the middle of the night and sailing to Crimea, coming within miles of the strategically vital Russian naval base of Sevastopol. This caused a flurry of social media activity as webcams from Odesa showed the two ships never left the port, meaning that someone had created false AIS tracks to trick OSINT users of MarineTraffic into believing NATO had violated Russia's security.

> **There is a kind of 'arms race' between OSINT analysts and the authors of disinformation.**

Stories in mainstream media outlets about this episode claimed Russia had spoofed AIS data; launched a GPS cyberattack; placed a nefarious AIS transmitter nearby; or interfered with the GPS. However, many of these misunderstood how AIS works and how it relates to open-source tracking websites as a form of socio-technical system. MarineTraffic.com makes it easy for volunteers to submit a data report so that anyone, anywhere in the world, can submit data. However, such reports are not verified, and just because something is displayed on the website does not mean it is happening in the physical world. Although many of the experts cited in the media discussed the technical sophistication of the systems involved, they missed the relatively easy-to-manipulate vulnerabilities to seed disinformation on them at time-sensitive moments.

The broader point, however, is that influential sources of disinformation in the contemporary information environment are not confined only to media and social media. Consequently, the OSINT community needs to widen their radar and toolkit for detecting potential vulnerabilities and exploits.
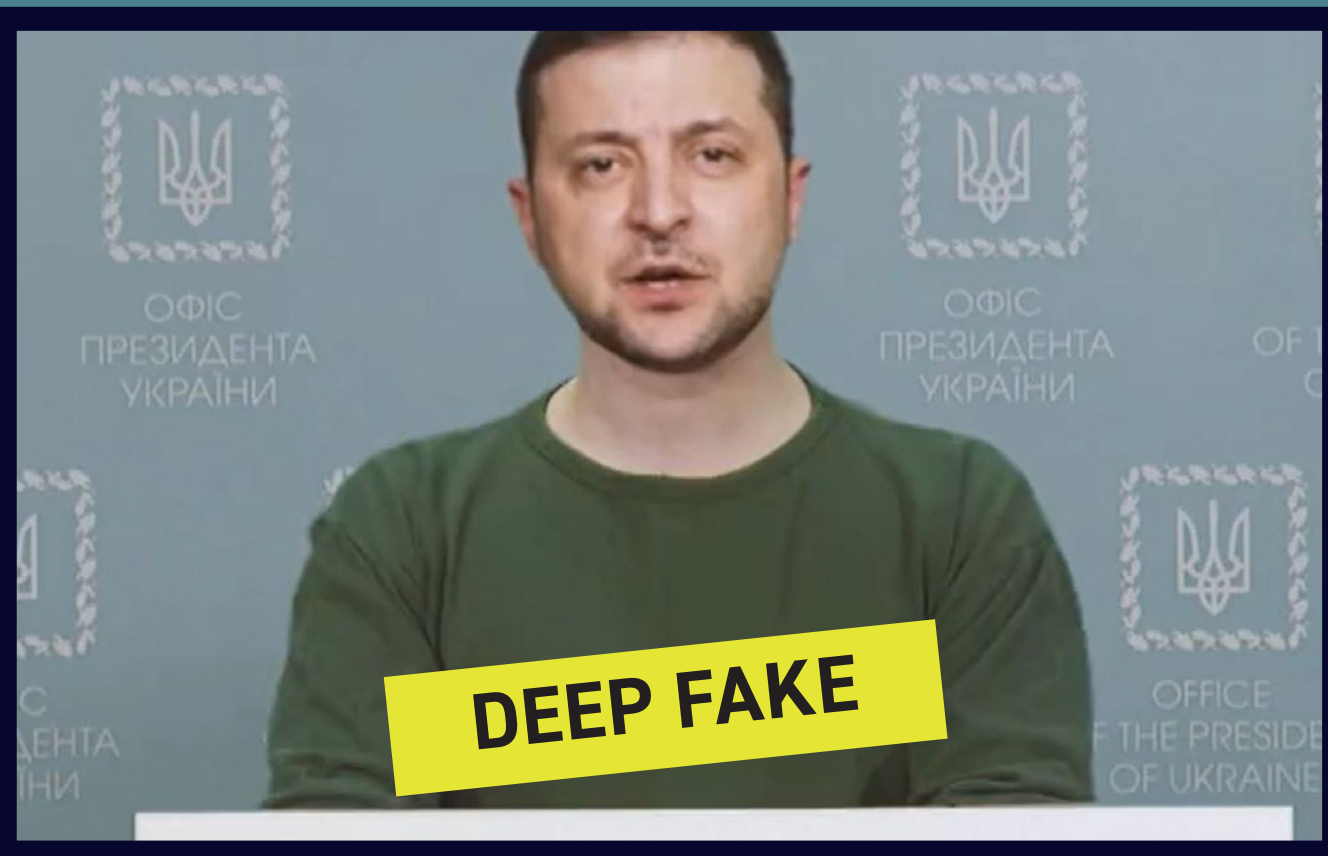
REAL

DEEP FAKE

Image credit: (top) © President Zelensky, Handout/Anadolu Agency via Getty Images
(bottom) a screenshot from the now deleted deep fake video of President Zelensky

ARTIFICIAL INTELLIGENCE AND DEEP FAKES

Following Russia's invasion, Ukrainian officials publicly warned that adversaries might be preparing a deepfake video of President Zelensky announcing his surrender. At the time, it was unclear if this was speculation or based on credible intelligence. However, less than two weeks later, a video was circulating on multiple social media platforms showing 'Zelensky' speaking directly to the camera. Although the manipulation was relatively unsophisticated and easy to spot, it is believed to be the first weaponised use of a deepfake during an armed conflict.

Social media platforms removed the video in violation of policies on the deceptive use of synthetic media, and Zelensky quickly debunked it. The early timing of its release, its central message, "lay down your arms and return to your families… I am going to do the same" was clearly intended to disorient and cause panic and doubt. It co-existed with disinformation coming from Russian officials that Zelensky had fled the country, contrary to Zelensky's own highly effective use of social media to broadcast 'proof of life' videos from the centre of Kyiv, the day after Russia attacked.

Deepfakes are at the cutting edge of artificial intelligence (AI) and machine learning algorithms can digitally forge a manipulated image of an individual using material sourced online. Also, in March last year, a Putin deepfake used clips from his televised Presidential address, adding new audio to make him appear to be surrendering to Ukraine. It was such poor quality and Putin's words so incongruous that audiences widely regarded it as satire, but it is certain that technological capability and expertise will rapidly advance to challenge human capacity to discern what is real and what is fake. For a lesson in how rapidly AI technologies evolve and become widely accessible and multi-purpose, look no further than ChatGPT. This large language model chatbot was only launched at the end of last year, but over 100 million users have queried it for many different purposes, some more nefarious than others.

In the hands of malign state actors, such AI-assisted technologies can create a high-volume stream of potent disinformation. An AI-assisted writing tool was recently used to generate misleading citations in a news website article about Russian opposition leader Alexei Navalny, for example. Automated text generation will facilitate the mass creation of social media accounts that look more authentic to users, whilst it appears inevitable that visual disinformation in the form of deepfakes will be deployed with other techniques of information warfare, such as hacking. The outcomes will exacerbate social tensions at critical moments of war or elections, damaging the credibility of its targets. Even a growing volume of poor-quality, more readily accessible media manipulation techniques ('shallowfakes' require only basic editing software) will erode public trust in news media.

> "In the hands of malign state actors, such AI-assisted technologies can create a high-volume stream of potent disinformation.

CONCLUSION

The methods via which disinformation is authored and amplified are rapidly evolving and adapting. There is understandable concern that new tools and technologies will enable false and misleading messaging to be produced at a pace and scale that will overwhelm our capacity for information defence. It is also worrying that increasing numbers of actors, both state and non-state, appear to be seeing information manipulation as a key tactic and technique for achieving digital influence in the information age. In 2024 there are important elections scheduled across the UK, US, Russia and the European Union, amongst others. It is vital and urgent to consider how OSINT methods can be re-tooled and 're-armed' against future disinformation threats, to mitigate or slow this advancement.

Helen Innes is a Research Fellow at Cardiff University Security Crime and Intelligence Innovation Institute. Her work identifying and analysing disinformation campaigns and foreign state information operations contributes to the Disinformation, Strategic Communications and Open Source Research Programme.

Andrew Dawson is a Research Associate at Cardiff University Security Crime and Intelligence Innovation Institute. His work spans topics such as Automated Facial Recognition, terrorism, and the Internet Research Agency. His recent work focuses on exploiting Open Source Intelligence for the Disinformation, Strategic Communications and Open Source Research Programme.

Martin Innes is a professor at Cardiff University and Co-Director of the Security, Crime and Intelligence Innovation Institute. His work on policing, counterterrorism, and disinformation has been internationally influential across the academic, policy and practice communities.

BETTINA ROTTWEILER & PAUL GILL

CONSPIRATORIAL THINKING AND FAR-RIGHT EXTREMIST ATTITUDES

Exploring the contexts that amplify or dampen the relationship between conspiratorial thinking and far-right extremist attitudes.

Over the last few years, we have seen a series of recent far-right terrorist attacks which demonstrate that extreme conspiratorial worldviews can mobilise individuals towards extremist violence. For instance, the Pittsburgh synagogue shooting in October 2018, the Christchurch terrorist attack, which left 51 people dead at a mosque and community centre in New Zealand in March 2019, the attack targeting Hispanic people at a Walmart in El Paso, Texas, in August 2019 as well as the more recent shooting in Buffalo, New York, which saw 10 Black people killed in a grocery store people in May 2022. These events were all heavily influenced and motivated by far-right conspiracy theories such as the 'great replacement' or 'white genocide'. This relationship has been further highlighted by the U.S. Capitol attack on January 6, 2021, which demonstrated an increasing synergy between extremist groups and QAnon adherents engaging side-by-side in anti-government violence.

Additionally, there has been an increasing convergence between far-right extremists, anti-vaxxers as well as conspiracy theorists within online spaces, which has led to threatening and violent behaviours towards politicians, health professionals and the media and thus, further demonstrates the radicalising effects of conspiracy beliefs.

We tested the effects of specific group and social-environmental influences on far-right extremist attitudes and we examined whether these relationships were amplified in the presence of conspiratorial mindsets.

In June 2019, we conducted a German nationally representative telephone survey with 1502 participants. This was achieved through a systematic and controlled approach of a multi-stratified probability sample (Random-Digit-Dialling) in the dual-frame mode (landline telephone-households and mobile phone users).

> **38.3%** of respondents showed agreement on average across all conspiracy mentality items.

We measured a wide range of different social-environmental and group-level factors, such as exposure to extremist peers, perceived threats to the ingroup and perceptions of group-based injustice. In addition, we asked each participant about the degree to which they agreed with:

1. Five generic themes that re-occur in different conspiracy theories

2. Six statements tapping into seven different dimensions:
   1. Support for a right-wing dictatorship
   2. Chauvinism
   3. Xenophobia
   4. Anti-Semitism
   5. Social Darwinism
   6. Islamophobia
   7. Downplaying the crimes of National Socialism

All items were measured on 7-point Likert scales, where '1' meant 'I strongly disagree' and '7' meant 'I strongly agree'. Responses of 'somewhat agree', 'agree' and 'strongly agree' were considered to show an 'agreement' with the statements. We further estimated mean scores (m), for each item and also calculated an average score for all individual scale items combined where appropriate (e.g., all items measuring far-right extremist attitudes), ranging from 1 – 7.

The following table presents sample of statements that we measured, together with the percentage who agreed and mean scores.

Conspiracy Mentality	
There are secret organisations that greatly influence political decisions	53% (m = 4.4)
Events which superficially seem to lack a connection are often the result of secret activities	36.1% (m = 3.7)
Government agencies closely monitor all citizens	36.4% (m = 3.7)
Far-Right Extremist Attitudes	
National Socialism also has its positives	9.9% (m = 1.9)
Jews bear some of the blame for their persecutions during the Third Reich	6.6% (m = 1.7)
Foreigners only come here to exploit our welfare state	23.1% (m = 3.0)

38.3% of respondents showed agreement on average across all conspiracy mentality statements (m = 4.3), whereas on average across all far-right extremist attitudes items 5.3% (m = 2.6) indicated support for far-right extremism.

We ran a series of moderation analyses, which determine whether the relationship between a predictor and an outcome depends on (i.e., is moderated by) the values of a third variable. A significant moderation effect indicates that the nature of the relationship between the predictor and the outcome changes, depending on the values of the moderating variable, e.g., when looking at low levels, average/mean levels, or high levels of the moderator.

In our study, we examined the moderating effects of different levels of conspiratorial beliefs (i.e., low, average, and high levels) on the nature of the relationship between the socio-environmental and group-level predictor variables and our outcome variable, far-right extremist attitudes. In other words, we tested whether varying levels of conspiracy thinking would potentially amplify or dampen the influence of:

1. Exposure to extremist peers,

2. Perceived threats to their ingroup, or

3. Perceptions of group-based injustice on far-right extremist attitudes.

We found that these three factors have particularly strong effects on support for far-right extremist attitudes amongst those individuals who also demonstrate high levels of conspiratorial mindsets. This also indicates that the effects of these three factors are all contingent on individuals' levels of conspiratorial thinking, whereby the harmful effects are most pronounced amongst those with 'strong conspiracy beliefs.'

Conversely, when conspiratorial thinking is low or average, there is a significantly reduced likelihood of support for far-right extremist ideologies. Importantly, low levels of generic conspiracy thinking are able to dampen the risk effects on far-right extremism, despite individuals being exposed to extremist peers, experiencing threats to the ingroup and demonstrating high levels of perceptions of group-based injustice.

The findings demonstrate that exposure to extremist peers, perceived threats to the ingroup, perceptions of group-based injustice and conspiratorial thinking all independently predict increased far-right extremist attitudes.

> **When conspiratorial thinking is low or average, there is a significantly reduced likelihood of support for far-right extremist ideologies.**

However, the strength of the relationships between the socio-environmental and group-level risk factors and far-right extremist attitudes depends upon individual levels of conspiratorial thinking.

Consistent approval of conspiracy mentality statements was around seven times more prevalent than consistent support for far-right extremist attitudes. This shows the former, by itself, is an insufficient explanation of the latter. Other factors must be at play.

Our results show that group- and social-environmental risk factors become particularly harmful when they co-occur with conspiratorial thinking - it is the interactive effect of these factors which significantly increases support for far-right extremist ideologies.

Yet, these findings also suggest that certain individual level factors (e.g., low levels of conspiracy thinking) can exert interactive protective effects by dampening, mitigating, or nullifying certain risk effects upon far-right extremist attitudes. From an intervention perspective, this may be particularly relevant when working with people 'at-risk' or those vulnerable to radicalisation by considering those factors which seem to play a functional role in amplifying or dampening existing vulnerabilities.

Collectively, the findings demonstrate the importance of acknowledging conditional and contextual factors to capture a much more nuanced picture of this complex relationship.

Dr Bettina Rottweiler is a Postdoctoral Researcher in the Department of Security and Crime Science at University College London. Her research examines the underlying risk and protective factors for different violent extremist outcomes for use in research and practice, with a specific focus on the functional role of conspiracy beliefs and violent misogyny within violent extremism.

Paul Gill is a Professor of Security and Crime Science at University College London. Prior to joining UCL, Professor Gill was a postdoctoral research fellow at the International Center for the Study of Terrorism at Pennsylvania State University. He has over 80 publications on the topic of terrorist behaviour.

Image credit: © *Becker1999* | Flickr

ZOEY REEVE

MISLEADING A GROUP TO INEFFECTIVENESS

To understand the effects of misleading information and distraction on group performance, Zoey Reeve and team monitored groups of people working together online to solve a murder mystery.

Disrupting criminal behaviour is an important endeavour in countering organised crime and terrorism. Tactics to achieve this often focus on resource availability (e.g., seizure of funds, taking down websites), or group dynamics (e.g., removing a leader, affecting communication channels). Questions are often asked about how effective a tactic is in absolute and relative terms, and whether its effectiveness is moderated by other factors. We've started to examine these questions within our lab, and repeatedly find that providing misleading information results in the worst group performance.

> **Disrupting criminal behaviour is an important endeavour in countering organised crime and terrorism.**

TO MISLEAD A GROUP

When we mislead a group, we provide them with information that directs their attention away from other, possibly more credible, sources. Such irrelevant information (which may manifest as misinformation or disinformation), often leads to mis-framing of a problem as it shapes the 'narrative hypotheses' that a group may use to solve a task. This can be especially problematic when it occurs early in task planning as the inclusion of incorrect information becomes embedded into an understanding of a particular problem or solution and results in poor judgements and decision-making, which can continue even after being corrected.

Although we often think about misinformation and/or disinformation when we consider misleading information, we might see it manifested in other ways, such as through tactics to distract. Distraction may eminate within a group through 'bad apples'. Bad apples freeload and shirk their own effort, or – and of relevance here – they might act to shift the effort of the group away from the task at hand and towards less relevant factors (e.g., intragroup conflict). Thus, the behaviour of bad apples may influence other group members towards aggression, defensiveness, and withdrawal, by disrupting relationships, whilst misleading information disrupts performance.

WHICH TACTIC IS MOST EFFECTIVE?

We wanted to understand the effects of false information and distraction on group performance. We formed 79 groups of up to five people to work together online for five weeks to solve a murder mystery. The groups were provided evidence over four weeks to identify the murderer in week five. We intervened in the activity of 48 groups in weeks 2 and 4: half of these groups received misleading information (a piece of evidence pointing towards a non-guilty character) and later a distraction (a set of chat prompts used by a confederate to detract focus away from the evidence and the task at hand), while the other half received the same interventions in reverse. The remaining 31 groups received no intervention.

We found that groups who experienced an intervention were much less likely to solve the task. In particular, those exposed to misleading information early on were more likely to fail at the task than those who we distracted early or those who received no intervention. One reason why misleading information negatively impacted group performance was because it reduced social cohesion. A reduction in social cohesion generally results in group members being less committed to the joint goal, and

> **Misleading groups is an effective tactic to disrupt criminal behaviour.**

less willing to be influenced by others or to voice disagreement when making joint decisions. One way that cohesion can manifest – and the way we measured it here – was through language. Cohesive groups tend to communicate equally across all members and in a positive way (e.g., agreement and positive emotion) and show more alignment in their linguistic styles.

Our results showed that linguistic alignment was lowest in those groups that received misleading information, which contrasted with the strong alignment we recorded in groups that were not exposed to an intervention.

Misleading groups is an effective tactic to disrupt criminal behaviour. Misled groups perform tasks less successfully than non-misled groups. However, to successfully mislead groups, disruptions need to occur early on in an interaction. Attempting to mislead groups later in a process is less impactful because the additional time promotes the development of group cohesion, and attention focused more on credible/accurate sources. However, our results suggest that not all disruptions are created equally. Exposure to misleading information appears to be more effective than distraction. Inaccurate information becomes encoded into thinking and the way in which solutions are shaped. It also has a strong impact on social cohesion. One implication of these results is that interventions that significantly reduce social cohesion are likely to be effective in derailing group performance.

Dr Zoey Reeve is a Senior Research Associate at the University of Lancaster, with CREST.

TOM BUCHANAN, ROTEM PERACH & DEBORAH HUSBANDS

WHY DO PEOPLE SHARE FALSE POLITICAL INFORMATION ONLINE?

People regularly encounter false political information on social media. Perhaps one in ten forward it on. Online misinformation spreads far and fast, with potential consequences for attitudes, beliefs, and actions.

Considerable effort has been invested in attempts to counter the spread of false information, with mixed success. In addition to fact-checking and debunking, psychological interventions have been developed. These include inoculation, gamified interventions, and approaches which aim to focus attention on accuracy. All of these approaches can work. However, there are issues with the scalability of some interventions, while others rely on cooperation from social media platforms. Furthermore, a recent re-analysis of data from gamified interventions suggests they may reduce trust in information overall, rather than enhance our ability to tell truth from falsehood. Re-analysis of data from interventions prompting people to consider accuracy suggests the approach is ineffective for politically conservative people.

Our work focuses on why people share false information. Some individuals share misinformation because they genuinely believe it to be true, while others knowingly share false content. Some personality types may be more likely to engage with false information than others. As motivations influence the effectiveness of interventions, it is useful to understand these in order to know if an intervention to help people recognise false information is likely to be effective for those who will share it anyway to achieve some desired outcome, or only be effective for those who believe the information to be true.

> **Debunking false information is unlikely to be effective for those who will share it anyway.**

MOTIVES FOR SHARING

Based on social media users' own accounts, we identified six distinct sets of motives for sharing political information and misinformation. Three sets — prosocial activism, awareness, and fighting false information — demonstrate a desire to 'make things better', benefiting other individuals and society as a whole. The other three sets reflect motives relating to attack or manipulation of others, political self-expression, and entertainment. Sharing misinformation can therefore be driven by destructive motives, but also be seen as a strategy to enhance social cohesion.

PROSOCIAL ACTIVISM

A desire to educate, inform, or mobilise other people in ways intended to benefit them or society. Sentiments about driving social change, critical thinking, morality, and political accountability, as well as informing people. Proactive use of social media to achieve political or social goals regarded as positive by the individual, and not involving tactics such as attacking others.

AWARENESS

These motives seemed to revolve around transparency or making people aware of information, and reflect 'good' reasons for sharing information. However, themes appear to be tinged with suspicion, and may be indicative of conspiracist ideation.

FIGHTING FALSE INFORMATION

Combating misinformation and minimising its harm, generally reflecting social responsibility in the political misinformation domain. Individuals endorsing these items might try to debunk false information (even if inadvertently spreading it further while doing so).

ATTACK OR MANIPULATION OF OTHERS

Cynical, antisocial, and manipulative use of social media. A desire to achieve one's own ends with a disregard for the truth or the welfare of others. Some of the motives dealt with self-enhancement. Others dealt with actively doing harm to others. Overall, these sentiments were either directly opposed to 'prosocial activism' motives, or treated as irrelevant.

POLITICAL SELF-EXPRESSION

Expression of political views and participation in political debate. People endorsing these motives want to talk about politics, not necessarily to bring about political change.

ENTERTAINMENT

A desire to entertain oneself or others, be funny, or alleviate boredom.

INDIVIDUAL DIFFERENCES

It is important to consider the characteristics of people who engage with false information online. Some research suggests specific personalities are more likely to share misinformation (for example, people who are politically conservative and have low levels of conscientiousness). It is also suggested that some interventions may only be effective for particular types of people, such as accuracy nudges only being effective for politically liberal individuals. However, research on personality and demographic variables has produced conflicting results, making it challenging to draw definitive conclusions.

Findings from several of our studies suggest that schizotypy may be important. Schizotypy is a set of characteristics associated with disordered thinking. It has multiple dimensions. Positive schizotypy is associated with suspicion, disordered perception, and belief in the paranormal. We have found that people with higher levels of positive schizotypy are more likely to report sharing false information. These findings are based on self-report data, and we need to extend this to evaluate behavioural evidence.

PRACTICAL IMPLICATIONS

While effective interventions have been developed, they may not be universally effective. For example, a truth-discernment protocol might work for those motivated by prosocial

> **Going 'all in' on one specific type of intervention may be unwise.**

activism, but it is unlikely to be effective for those who share political information with the intention of attacking or manipulating others. Additionally, certain people may be more vulnerable to misinformation, or resistant to particular interventions. This means going 'all in' on one specific type of intervention may be unwise. Further research is needed on individual characteristics that influence engagement with misinformation. This should be considered within the wider picture of general vulnerability to online influence. Finally, more work is needed to evaluate the actual effects of exposure to false information online.

Tom Buchanan is a Professor of Psychology at the University of Westminster. Dr Rotem Perach is a Research Fellow at the University of Westminster. Dr Deborah Husbands is a Reader in Psychology at the University of Westminster.

WILLIAM DANCE

ADDRESSING ALGORITHMS IN DISINFORMATION

A look at how people discuss false content online and how exploring social media discourses can help strengthen policy responses.

In February 2024, the European Union's 'Digital Services Act' (DSA) will come into effect. The DSA will enforce a standard of transparency on very large social media platforms, obliging them to lay out how their sophisticated, proprietary content recommendation algorithms work. The act is in response to years of algorithmically fuelled disinformation that has undermined public trust and led to real-world harms (Jolley & Paterson, 2020; Wardle & Singerman, 2021).

Algorithms and the spread of disinformation are inexorably linked. Algorithmic recommender systems that suggest new content to users may serve as a vector between disinformation producers and social media users, potentially delivering false and harmful content. Understanding these systems, their effects, and public perceptions of algorithms is vital to forming legislation that responds to such threats.

PUBLIC PERCEPTIONS OF DISINFORMATION

My research uses corpus linguistic approaches to study the replication and reception of online disinformation on social media. I focus on how, linguistically, people share false content online and how ideas on the internet spread from their inception until they cease to exist. This involves exploring metacommentary around disinformation, or more simply looking at how people talk about disinformation itself.

Understanding how the public talk about important topics is a tried-and-tested method for understanding them with greater nuance, whether it's discourses of Islam (Baker et al., 2013), discussions of vaccination (Coltman-Patel et al., 2022), or exploring hate speech online (Hardaker & McGlashan, 2016). Disinformation poses a security threat by clouding decision-making at both individual and national levels. Understanding how the public perceives disinformation is crucial to mitigating its effects.

Algorithmic disinformation is a complex issue that requires an equally complex solution, combining regulation, policy, education, and fact-checking. But what if the public do not always see it as a concern? In an analysis of almost 40,000

tweets* spanning the first six months of 2022 containing the words disinformation, misinformation, or fake news, the word 'algorithm' is mentioned in just 24 tweets. To put this in perspective, the word 'dog'/'dogs', something unrelated to the topic at hand, is mentioned in 31 tweets. That is to say, people discuss dogs more often than they do algorithms in relation to disinformation, misinformation, and fake news in the dataset.

> **...people discuss dogs more often than they do algorithms in relation to disinformation, misinformation, and fake news...**

This has implications for how we tackle algorithmic disinformation online because if awareness is low, policy responses such as the DSA may be viewed as disproportionate in scale in terms of public perceptions of the issue. Algorithms are fundamental to social media and the spread of disinformation online. While a lack of explicit mention does not imply a complete lack of knowledge, there seems to be an awareness gap. This data offers a snapshot of discussions, and given extensive policy responses to disinformation, it is vital to learn from these findings.

When the public does discuss disinformation, they are keenly aware of its dangers. Online discussions specifically highlight the threat to democracy caused by disinformation, how it infringes on human rights, and its disproportionate impact on issues such as reproductive healthcare. Throughout, disinformation is framed as an enemy, something we should fight and combat. There is, however, a paradox here. Research has shown that simply discussing disinformation and its negative effects can

affect key metrics such as trust and cynicism (Jones-Jang et al., 2020; Vaccari & Chadwick, 2020). Therefore, when addressing disinformation, we need to be aware that overexposure to the topic can do more harm than good.

INFORMING POLICY RESPONSES

The public is aware of disinformation's harmful potential to threaten civil liberties and impact our institutions but they are not necessarily familiar with the nuances of how disinformation spreads through technologies such as algorithms. Responses to disinformation should prioritise the human aspect, and the technical and social aspects of disinformation should not be seen as separate but rather as interconnected elements. Examining people's real-world concerns in natural settings helps us grasp

what troubles them and how changes in our online information environments can tackle the genuine worries related to the dissemination of disinformation.

Further, it is crucial to ground policy responses to security threats in real-world situations for an effective approach. Policies that address the public's genuine concerns are more likely to garner public support and foster positive change, helping to reduce the impact of disinformation. This includes addressing health threats such as disinformation that rejects conventional medicine and responding to information operations that use disinformation as a medium to undermine democracy. The individuals most at risk from disinformation are the public themselves, and it is their concerns that should guide our response to disinformation.

> **...when addressing disinformation, we need to be aware that overexposure to the topic can do more harm than good.**

William Dance is a PhD student and Senior Research Associate in the ESRC Centre for Corpus Approaches to Social Science at Lancaster University. His research combines historical approaches to studying language with the analysis of contemporary social media datasets to explore the development of disinformation over centuries.

*Twitter is now called X, and tweets are now called posts.

ISABELLA ORPEN

USING FUZZY SET QUALITATIVE COMPARATIVE ANALYSIS TO EXAMINE HETEROGENEITY IN CONSPIRACY BELIEVERS

The plethora of security ramifications of conspiracy beliefs range from public health endangerment to violent attacks on democratic institutions. Understanding diverse types of conspiracy believers is crucial to better understanding and mitigating the potential risks.

Scholars repeatedly note the radicalising power of conspiracy ideologies. Storming of the Capitol and eruptions of extremism during the coronavirus pandemic illustrate the security implications of strongly held conspiracy beliefs. Even more casually held conspiracy beliefs, such as mistrust of vaccines and information surrounding coronavirus, can severely impact following government guidance and therefore public health. The coronavirus pandemic and the global rise of populist conspiracy theories show that the appetite toward conspiracy ideation is more commonplace than previously thought.

Conspiracy beliefs often run counter to official narratives and centre around a group of malicious conspirators and their hidden involvement in seemingly unrelated events. Conspiracy theory belief is a term often used in research to refer to both conspiracy mentality (i.e., the propensity toward conspiracy beliefs) and belief in specific conspiracy theories (e.g., anti-vaccination myths). This understanding of conspiracy beliefs can encompass a wide array of beliefs.

This definition makes no judgement on the veracity or morality of such beliefs. There is ongoing debate in academic literature and the media regarding whether conspiracy beliefs are solely harmful and divisive for society or if critical examination of official narratives can have benefits. This debate is wrongly based on the assumption that conspiracy beliefs are uniform (homogenous) and that all conspiracy believers are alike. More recent research has called for greater understanding of the diversity (heterogeneity) of conspiracy beliefs and believers.

One way we can examine the heterogeneity of conspiracy beliefs and believers is through Fuzzy set Qualitative Comparative Analysis (fsQCA). This method combines the contextual richness of qualitative case-based analysis and the rigour of quantitative analysis. Adding a degree of fuzziness to the variables' membership means a further level of nuance can be achieved.

> **Adding a degree of fuzziness to the variables' membership means a further level of nuance can be achieved.**

1. CONSPIRACIES ARE FUZZY BY NATURE

Conspiracy scholars disagree over whether conspiracy beliefs are a product of an extreme minority or a general human tendency. Recognition of different levels of conspiracy belief can help to reconcile these two disparate schools of thought. The survey data collected by the research team at SCIII, Cardiff, shows that almost three-quarters of British respondents exhibit some level of conspiracy mentality and that a third agree with anti-vaccination myths and do not trust the official mortality rates reported for coronavirus. Within this group of 'believers' we see a difference in the strength of beliefs, with some showing a 'strong' belief across all indicators and others showing mixed or casual belief across the indicators. FsQCA is well-suited to address the nuanced nature of conspiracy beliefs, which often exist in a grey area between fact and fiction. It offers a framework where complex concepts can be measured on a continuum rather than relying on binary distinctions. By incorporating thresholds, fsQCA helps determine the strength of beliefs and identifies the point at which they transition from weak to strong.

2. COMPLEX COMBINATIONS OF CONDITIONS

Research has identified psychological traits (e.g., paranoia) and social beliefs (e.g., patriotism) as key 'conditions' in conspiracy beliefs. However, these conditions are almost exclusively identified through regression-based net effect models. Attempts to isolate each condition's unique impact on conspiracy beliefs fail to account for their interconnectedness.

FsQCA allows for conditions to have varied effects depending on their configuration with other conditions. This nuance is hindered by methods that only consider the net effects of factors on conspiracy beliefs, as they aim to isolate and identify the impact of individual conditions instead of embracing their contextual complexities.

3. DIVERSITY IN PATHWAYS TO STRONG BELIEF

Heterogeneity in conspiracy beliefs goes beyond the intensity of belief. The origins of these beliefs are also crucial to consider. FsQCA allows multiple 'pathways' (configurations of conditions) to lead to the same outcome. Therefore, heterogeneity among conspiracy believers can be understood in terms of both the strength of their belief and the different pathways that led them there. For those with a strong conspiracy mentality and strong coronavirus conspiracy beliefs, multiple configurations can result in the same outcome. This underlines the value of using configurational methods (such as fsQCA), as there is not a single pathway to strong conspiracy beliefs.

4. EXAMINING THE ABSENCE OF CONSPIRACY BELIEFS

FsQCA considers both the presence and absence of an outcome, such as strong conspiracy beliefs. Unlike other techniques, fsQCA recognises that the presence of certain conditions associated with the outcome does not guarantee that their absence will lead to a negated outcome.

This understanding is pivotal in addressing the risks associated with conspiracy beliefs, as research often focuses on identifying risk factors that can be used to prevent harmful behaviours. FsQCA highlights that practitioners should identify not only risk factors but also protective factors, while explaining why the two may not perfectly align with each other.

This article has shown the adoption of new analytical methods to examine the complexity and diversity of conspiracy beliefs and believers. FsQCA presents an interesting methodological solution to the intricacies of conspiracy beliefs. Broadening our understanding of different types of conspiracy beliefs can help to understand at what point they pose a security risk and how this can be mitigated.

Isabella Orpen is a Research Assistant at the Security Crime and Intelligence Innovation Institute at Cardiff University. She is currently completing her PhD looking at understanding the heterogeneity of conspiracy believers. The results of applying fsQCA will form part of her PhD.

Image credit: © *Alina Kolyuka* | stock.adobe.com

GRACE MCKENZIE

HIDING IN PLAIN SITE

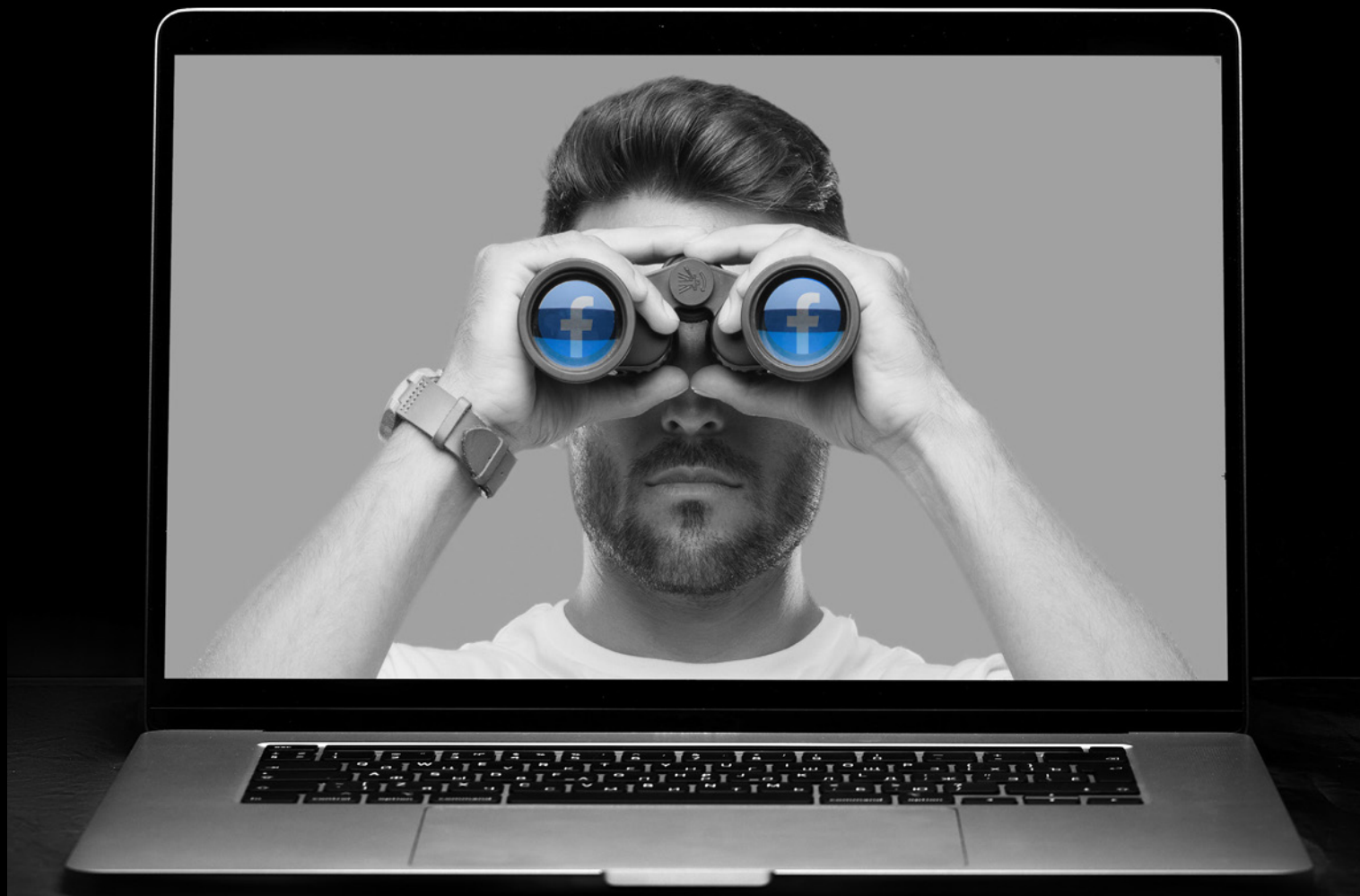## How do you decide that a social media profile is fake? What happens if your judgement is wrong?

The introduction of social media has had a massive, almost incomprehensible, impact on society and the way in which we communicate. With the good of these advances in modern life, comes the bad. Not only are there the 'Dark Web' and the 'Deep Web', where criminal and malicious transactions occur, but even everyday social media brings a plethora of danger in the form of scams, misinformation, fake news, and fake profiles, to name but a few. But how big is the problem? And what does this mean for social media users?

> **4-5% of active accounts on Facebook are fake.**

THE DANGERS OF SOCIAL MEDIA

As of January 2023, 4.76 billion (59.4%) of the world's population use social media (Statista, 2023). Facebook, the largest platform worldwide with 2.59 billion users (Statista, 2023), estimates that 4-5% of active accounts on Facebook are fake (Meta, 2023); which equates to between 103.6 million - 129.5 million accounts. Meta is actively trying to identify and remove fake accounts from their platforms. Within their open access 'transparency centre' (Meta, 2023), Meta reported that during the latter quarter of 2022, 1.3 billion fake Facebook accounts were identified and removed. For the first quarter of 2023, this number reduced dramatically to 426 million, the first time in over four years that number had dropped below the one billion mark. Why? Meta reported that this drop was expected, as the nature of the platform is 'highly adversarial'. However, could this be due to the huge developments in AI technology? Or faltering algorithms? More importantly, what happens when the computers cannot detect the accounts? Can humans detect them?

Such questions do not yet have a definitive answer. The consensus of current research in AI technology, specifically machine learning algorithms and social media bots, is that the detection accuracy rate is over 90% (Kudugunta & Ferrara, 2018; Chavoshi, Hamooni, & Mueen, 2016). Profiles that fall through the net are obviously problematic as they continue to spread misinformation and pose a threat to platform users through scamming and catfishing. So, how can users protect themselves from fake profiles when the platforms themselves cannot?

To answer this question, my research focuses on fake profile detection from a psychological perspective – namely humans' ability to detect deception in the online space – and human judgement accuracy. To assess humans' ability at identifying fake profiles, a collection of real and fake profiles are shown to participants and their task is to judge which are authentic. The results across five studies show that participants consistently judge real profiles more accurately than fake profiles, with participants achieving an average of 79 – 86% judgement accuracy for real profiles but only 13 – 54% judgement accuracy for fake profiles. People's ability to accurately judge a fake profile seems to improve the further away the profile gets from what we may consider as typical or 'normal'.

Of the 924 participants that have been tested, zero participants accurately judged all the profiles they were shown. The average accuracy score when shown a random selection of real and fake profiles was at 50%. This supports the well-cited meta-analysis of Bond and DePaulo (2006), which shows that humans' accuracy at deception detection is 54%. Interestingly, this result held even when the time taken to decide was varied (time constraint vs unlimited time) and when viewing profiles from a different culture.

To understand the specific areas of the profile that may influence the decision-making process, participants were instructed to click on the specific areas of the profile that they relied upon to make their judgement. Consistently across all five studies, participants relied most on the images on the profile, specifically the 'profile picture' or 'cover photo', when judging. This was the case regardless of whether they were reviewing a real or a fake profile. Participants also relied heavily on the content of the posts on the profile, but not to the same extent as the images. Contrastingly, areas such as the 'Intro' section containing information such as the person's location, school/university, job, relationship status etc., and the numbers of likes/comments on each post were relied upon much less, if at all.

FOOD FOR THOUGHT

Historically, people used to guard against admittance to secure areas with a verbal challenge of 'friend or foe?', with approved entry coming via a pre-arranged password. Now those secure areas are our virtual, online lives, and the challengers come in the form of fake profiles. It may seem that with failings in both software driven responses and natural human judgement error that we are no further forward. Hopefully, with the ever-expanding developments within AI technology, there is potential for the creation of a programme with an element of trained human oversight that can work towards greater fake profile detection accuracy rates. But for now, it seems that with all that modern technology has to offer, we're often left no better than the generations that preceded us. The age-old question remains in need of an answer: friend or foe?

> **People's ability to accurately judge a fake profile seems to improve the further away the profile gets from what we may consider as typical or 'normal'.**

Grace McKenzie is a final year PhD Psychology researcher at Lancaster University. Her thesis investigates human judgement of online deception. She is affiliated with CREST via her supervisors Professor Stacey Conchie and Professor Paul Taylor.

Image credit: © *Krakenimages* | stock.adobe.com

DIMITRI PAVLOUNIS & KELSEY DAVIS

WHAT KIND OF DIGITAL MEDIA LITERACY?:

BUILDING STUDENT RESILIENCE TO MISINFORMATION THROUGH EVIDENCE-BASED APPROACHES

Digital media literacy is often proposed as a solution to misinformation, but while some methods have been shown to be effective, many widely taught approaches can potentially cause harm.

It is a common refrain that students need to be taught digital media literacy in school to build resilience against misinformation. While this may be an uncontroversial statement in theory, it is ultimately meaningless when unmoored from specific practices. Indeed, the problem is not that students are not taught digital media literacy but rather that many of them are taught an assortment of outdated and untested methods that can leave them more vulnerable to misinformation and less trusting of high-quality information.

In Canada, for instance, the ability to assess online sources is a curriculum standard in every province and territory. However, in a Canada-wide study of over 2,300 students in grades 9-12, we found that students lacked the fundamental skills required to evaluate information effectively.

In perhaps the most striking example, students were shown a website from a group that presents itself as a medical research organisation but is actually a fringe anti-LGBTQI+ group that has been designated a hate group by the Southern Poverty Law Center (SPLC). They were given access to the internet and asked to rate the site's reliability as a source of medical information and justify their answer in an open response box. Sixty per cent of students rated the site as reliable, and only 6% determined the website's agenda.

ACCOUNTING FOR STUDENTS' POOR PERFORMANCE

To inform their decision, students consistently scoured the page or post looking for superficial signals of authority: they assessed whether the site looked 'professional,' they searched for typos, they checked to see if the URL was a .org or a .com (believing, incorrectly, that .orgs all belong to reputable organisations), and they read what the source said about itself on its own 'About' page. They tried to apply critical thinking skills to assess the credibility of the website's claims, but they did not have the necessary subject matter knowledge to make informed judgments. These strategies constantly led them astray.

Students did not devise these strategies from nothing. As researchers at the Stanford History Education Group (SHEG) have observed, many students were simply applying what they had been taught under the guise of digital literacy. These ineffective strategies are often packaged in checklists for educators to use to teach students how to evaluate information online. Perhaps the most well-known, the CRAAP Test, can trace its lineage to tools developed in the 1990s to help librarians decide which print material to buy. It is widely used to teach digital literacy despite never having been properly evaluated in that context.

> **A mounting body of evidence shows that students' ability to evaluate online information improves dramatically if they are taught the skills of lateral reading.**

LATERAL READING AND CONTEXT SEEKING

The good news is that a mounting body of evidence shows that students' ability to evaluate online information improves dramatically if they are taught the skills of lateral reading, a term coined by SHEG to describe the process of opening a new tab and conducting quick searches to learn more about a source or claim. Instead of asking students to look closely at online content for clues about its reliability, lateral reading empowers students to use the affordances of the web to gain important context before engaging with the content further.

In our own study, the same 2,300 students from across Canada who initially fared so poorly demonstrated remarkable gains after completing our lateral reading program. Prior to instruction, students showed evidence of lateral reading just 11% of the time. One week following instruction, students read laterally 59% of the time, and the quality of their assessments improved dramatically as a result. Students who trusted the aforementioned anti-LGBTQI+ site simply because it had a .org domain now conducted searches to learn more about who is behind a site before engaging with it. Students who once dismissed credible news stories simply because they contained typos now cross-referenced claims with professional media sources before deciding whether to believe them.

> **Lateral reading alone will not solve the problem of misinformation.**

Lateral reading alone will not solve the problem of misinformation. It cannot account for the myriad cognitive biases people bring to the information they consume and share. It also cannot address the technical and infrastructural issues that enable the global spread of misinformation. But it can begin to address a foundational skills deficit that makes people vulnerable to false and misleading information.

WE KNOW LATERAL READING WORKS, BUT THERE IS MUCH TO BE DONE

We know lateral reading is effective, but these skills are not spreading at the necessary speed or scale. Organisations like our own and SHEG provide evidence-based resources for educators, yet many resource portals for educators still provide lists of outdated tools. Major public awareness campaigns continue to promote ineffective strategies outside of educational settings.

Platitudes about the need for digital media literacy, in general, are not enough. What is required is a whole-of-society approach to communicate evidence-based best practices and to support the continued evaluation of interventions within different local and global contexts. We do not just need 'more' digital media literacy; rather, we need to be more intentional about precisely what kind of digital media literacy we need to meet the moment.

Dimitri Pavlounis is the Director of Research and Kelsey Davis is the Digital Literacy Program Manager at CIVIX. CIVIX is a Canadian civic education charity dedicated to building the habits of active and informed citizenship among school-aged youth through experiential learning programs.

RICKY GREEN, IMANE KHAOUJA, DANIEL TORIBIO-FLÓREZ & KAREN M. DOUGLAS

CONSPIRACY THEORIES:
THEIR PROPAGATION AND LINKS TO POLITICAL VIOLENCE

Conspiracy theories can exploit societal insecurities, be propagated relatively easily, and incite political violence. Proactive strategies are essential for mitigating their influence and preventing their potential consequences.

THE PSYCHOLOGY OF CONSPIRACY THEORIES

A conspiracy theory is a belief that two or more actors have coordinated in secret to achieve an outcome and that their conspiracy is of public interest but not public knowledge. Research suggests that people are drawn to conspiracy theories in an (often unconscious) attempt to satisfy unmet psychological needs, such as the need to feel secure and in control of one's life. For example, people who feel politically powerless find conspiracy theories particularly appealing. There is no evidence that this helps, however. Conspiracy theories increase feelings of existential insecurity, making people more prone to finding other conspiracy theories appealing and falling down "rabbit holes" that are difficult to escape.

> **People who feel politically powerless find conspiracy theories particularly appealing.**

THE PROPAGATION OF CONSPIRACY THEORIES

The rise of social media has made it easier for people to spread false information and conspiracy theories. Research has also shown that conspiracy and scientific information spread online differently. In one study, "conspiracy information was found to propagate deeper and be more viral than science information".

Artificial Intelligence (AI) applications also contribute to generating conspiracy theories. The emergence of generative models like ChatGPT has made it simple to create human-like texts. These models can also be leveraged to create deepfakes of political leaders by adapting their video, audio, and pictures.

AI-generated content, including deepfake videos, is becoming increasingly challenging to differentiate from authentic human-created information and videos.

AI-generated text is even perceived as more credible in some cases. This is because it effectively uses emotionally compelling language (see Figure 1) that captures readers' attention, motivating them to share it. Additionally, these models are capable of quickly generating high-volume text, which can create the illusion that uncommon opinions are actually more widespread.

✦ Default (GPT-3.5)

RG You will help me amend a statement by giving it an emotional amplification. It needs to fit into a tweet, so no more than 280 characters. Here is the statement: "The 2020 United States presidential election was rigged. Trump was robbed."

"The 2020 United States presedential election was undeniably, heart-wrenchingly rigged. Trump, a true patriot, was shamelessly robbed of the people's voice and trust. 😔🙏 #ElectionInjustice #TruthMatters"

Figure 1. An example of a political conspiracy theory with AI-generated emotional amplification. The emotional amplification of "heart-wrenchingly rigged" conveys a strong sense of emotional pain and turmoil and "Trump, a true patriot" adds a layer of emotional attachment and support for him. Note. This response was generated on September 5, 2023, using GPT-3.5, which is based on OpenAI's GPT-3 architecture.

LINKS TO POLITICAL VIOLENCE

There are important societal consequences of conspiracy theories, such as decreased intentions to engage in climate-change behaviours. Conspiracy theories have also been linked to negative experiences in interpersonal relationships. Furthermore, conspiracy theories were implicated in destabilising events such as the storming of the US Capitol on January 6 2021. More recently, and closer to home, German police arrested 25 people suspected of being part of a far-right terror cell, linked to the Reichsbürger ("Citizens of the Reich") movement. This political movement endorses conspiracy theories that portray the current Federal Republic of Germany as an illegitimate "deep state" that operates against the "still existent" German Reich.

Psychological research backs up these anecdotal accounts of the link between conspiracy theories and political violence:

- People who hold extreme political views appear more likely to believe in conspiracy theories, such as the belief that the government is controlled by a secret cabal of elites.

- Radical violent extremist groups often use conspiracy theories to justify their violence, demonise their enemies, and create a sense of urgency among their followers.

- Conspiracy theories can be used to radicalise people, by providing a sense of meaning and purpose to those who are feeling lost and powerless.

- People who believe in conspiracy theories are also more in favour of using political violence to achieve their goals.

- Conspiracy theories have the potential to add fuel to existing conflicts between groups.

MITIGATING THE RISKS

Ways to mitigate these risks require a collective effort from researchers, policymakers, technology companies, and the general public.

AI can be used to detect and counter the spread of conspiracy theories. For example, one AI model is now able to detect anti-vaccination and white genocide conspiracy theories on social media with 96% and 83% accuracy, respectively. These models could be used to automatically alert human moderators to conspiracy theories being shared on their platforms. Further, despite AI companies' current policies against creating conspiracy theories, it is still very easy to do so (see Figure 1). Therefore, AI companies should be urged to create stronger policies to monitor and handle the creation of conspiracy content.

Psychological inoculation is a technique used to reduce susceptibility to conspiracy theories. Like traditional vaccines, psychological inoculation involves exposing people to a weakened form of a conspiracy theory to decrease susceptibility to them in the future. A field study conducted on YouTube used brief videos to inoculate people against commonly used manipulation techniques (e.g., the use of emotionally manipulative language) which improved their recognition of these techniques and subsequent truth discernment. In a similar fashion, the Bad News Game exposes people to the tactics used by others who spread conspiracy theories, by having them play a game to amass followers using the same tactics. These techniques, also known as 'prebunking', are effective and could be employed through social media campaigns or government programs and other initiatives. A practical guide to prebunking misinformation can be found at: *bit.ly/prebunking-guide* (see Van Der Linden's article on p. 26 for more on this guide).

We need to understand more about whether addressing people's psychological needs or improving their analytical thinking skills can reduce the appeal of conspiracy theories. Tentative work in this area suggests an analytical mindset and critical thinking skills are the most effective means of challenging conspiracy beliefs.

SANDER VAN DER LINDEN

A PSYCHOLOGICAL VACCINE AGAINST MISINFORMATION

This article is an edited excerpt from Sander's book, 'Foolproof: Why We Fall for Misinformation and How to Build Immunity'.

Our team (the University of Cambridge's Social Decision-Making Lab) headed off to New York to collaborate with Google Jigsaw, Google's technology incubator. We met up with Beth Goldberg, Director of Research & Development and together with our colleague Stephan Lewandowsky, we started to have regular conversations with Google about how to inoculate people against extremism on social media.

Just like a vaccine introduces your body to a weakened version of a harmful virus, it turns out that the mind can be inoculated against harmful misinformation by exposing people to—and persuasively refuting—weakened doses of misinformation. The process of psychological inoculation works by: (a) forewarning people of an impending manipulation attempt, and (b) arming people in advance with the arguments and cognitive tools they need to counter-argue and resist exposure to persuasive misinformation (known as a 'prebunk'). Or, as one BBC journalist writing about our research put it: "Like Han Solo, you shoot first."

> **The mind can be inoculated against harmful misinformation by exposing people to—and persuasively refuting—weakened doses of misinformation.**

Beth was very interested in scaling our inoculation approach en masse via YouTube (owned by Google). One of the common techniques Beth identified is the use of 'false dichotomies'. A false dichotomy is a manipulation technique designed to make you think that you only have two options to choose between, while in reality, there are many more. Because YouTube doesn't really deal in headlines or social media posts, the issue here is that these more subtle rhetorical techniques are often being used — persuasively — by political Guru's in YouTube videos. From rants that spread fake news about Covid-19 and climate change to attempts to recruit people into QAnon and ISIS.

For example, one ISIS recruitment video explicitly aimed at Western Muslims was titled, "There is no life without Jihad" — a clear example of a false dichotomy: either you join 'jihad' or you cannot lead a meaningful life.

To produce the vaccine, we needed to synthesise weakened doses, so we started to make our own animated videos. The videos follow the inoculation format closely and start with an immediate warning that you (the viewer) might be targeted with an attempt to manipulate your opinion. We then show people how to spot and refute misinformation that explicitly makes use of these techniques by exposing them to a series of weakened examples (the microdose) so that people can easily identify and resist them in the future.

For example, in the video that inoculated people against the 'false dichotomy' technique, we pulled material from Star Wars III – Revenge of the Sith. We show the climactic confrontation between Anakin Skywalker, soon to become Darth Vader, and his mentor, Obi-Wan Kenobi. Obi-Wan says, "My allegiance is to the Republic, to Democracy!" to which Skywalker replies: "If you're not with me, then you're my enemy." This is clearly a false dichotomy. We explain to the viewer that Obi-Wan is simply trying to prevent Anakin from joining the dark side; just because he disagrees with Anakin doesn't automatically make them enemies. Obi-Wan points out the fallacy in his reply: "Only a Sith deals in absolutes."

We ran several large randomised controlled trials where we either exposed people to one of our short videos or a control video about 'freezer burn'. We then asked people how manipulative they found a series of arguments and how willing they were to share them with others. An example of the false dichotomy quiz would be the following post: "Why give illegal immigrants access to social services? Why should we help illegals when we should be helping homeless Americans instead?"

In the experiment, we asked people to rate many such misleading items in the hope that their ability to discern manipulative from non-manipulative content would improve.

This is exactly what we found. Unlike the control group, the inoculated groups became much better at identifying which posts contained a specific manipulation strategy and were

1. Emotional warning

Users are alerted that there are impending "attacks" to manipulate them.

2. Refute the attack

This looks like...

☑ Scapegoating: singling out a person or group for unmerited blame

Users are equipped to spot and refute a manipulative message.

3. Micro-dose

Example:

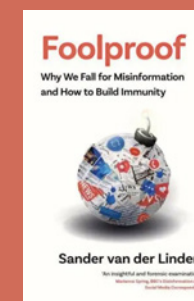Users see example(s) of manipulative message to identify it in the future.

subsequently less likely to want to share this type of content with others in their social network. Beth's idea was that we could implement and scale these videos on YouTube by inserting them in the 'non-skippable ad space' (you know, when you're trying to watch a video on YouTube and you get stuck watching an annoying ad that you can't skip? That's where our inoculation video would be placed).

We leveraged YouTube's ad platform to upload and target millions of U.S. users who are known to watch political content with either our inoculation or the control videos. Beth then got YouTube to agree to allow us to customise their 'brand lift' survey (which usually polls people on whether they recognise a brand) for a scientific experiment. Within twenty-four hours, they would be presented with a quiz in the ad space evaluating their ability to spot the misinformation technique they had been inoculated against (for example, the use of false dichotomies, emotional manipulation, scapegoating etc). We were able to reach about 5 million 'impressions' (views) with a single campaign. After watching the 90-second inoculation video, we boosted people's ability to spot misleading content by about 5–10 per cent. That might not seem much at the individual level, but this is in a realistic setting for a single dose of a short video clip that can be scaled across potentially hundreds of millions of people.

> **Can enough people be inoculated to achieve psychological herd immunity?**

Of course, the work is far from finished. There is much to explore about how applying the principles of psychological inoculation can empower individuals and policymakers to address societal challenges effectively. For example, we discovered that the vaccine wears off over time so campaigns ideally need to feature "booster shots" and feedback to enhance longer-term learning. In Google's latest prebunking campaigns—which reached the majority of social media users in Poland, Czechia, and Slovakia—they uncovered substantial cultural variation in the effectiveness of inoculation, suggesting that careful local tailoring is important. The ultimate question, of course, is whether enough people can be inoculated to achieve psychological herd immunity. Only then will misinformation no longer have a chance to spread. This will likely necessitate integrating inoculation against misinformation in our national educational curricula.

Foolproof
Why We Fall for Misinformation and How to Build Immunity

Sander van der Linden

ULLRICH ECKER, TOBY PRIKE & LI QIAN TAY

PSYCHOLOGICAL INTERVENTIONS TO COMBAT MISINFORMATION

Exploring misinformation's impact, resistance to correction, and the psychological strategies for debunking false information to create a healthier information environment.

FIRST THINGS FIRST: IS MISINFORMATION A PROBLEM? (YES, IT IS.)

Misinformation has become a buzzword, and many see the proliferation of misinformation and its potential impacts as an issue of substantial contemporary concern. We believe that, by and large, these concerns are justified. However, some argue that misinformation is only a) a small fragment of consumed information, b) a symptom rather than a cause of problems, c) has modest behavioural effects, and d) is nothing new. We disagree with these minimising arguments for several reasons:

a. While it is true that easily and objectively identifiable misinformation (e.g., 'fake news' headlines) makes up only a fraction of people's information diet, focusing on this subset of misinformation ignores all other types, including subtle misrepresentations and systematic distortions.

b. Broader societal issues and trends (e.g., social inequality and disenfranchisement; economic uncertainties; low trust in institutions) have likely causally contributed to enhanced misinformation spread and susceptibility. However, just because something is causally influenced by other factors does not mean it cannot have causal impacts of its own. For instance, there is evidence that misinformation has causally contributed to COVID-19 and MMR vaccine hesitancy, disregard for public-health advice, persecution of minorities, and the 2021 storming of the U.S. Capitol.

c. Measuring the impact of misinformation on behaviour is challenging due to its heterogeneity and the likelihood of being negligible or absent in certain cases (e.g., one-time exposure; low-plausibility misinformation; inconsequential topics). However, even small behavioural impacts can be meaningful at scale. Additionally, these impacts are not always direct; misinformation can indirectly shape people's views and choices by influencing mainstream media, public discourse, and policy-making in political debates. Moreover, there are likely additional ripple effects, such as diminishing institutional trust, which can further impact behaviour in distinct ways.

d. People providing false and misleading information is obviously not a new phenomenon. However, the fact that misinformation has long been present does not mean that it is no longer a concern. The misinformation problem has been exacerbated by rapid changes to the contemporary information environment. This is characterised by a growing reliance on the internet and social media as a primary source of information, unprecedented concentration of mainstream-media ownership, and the advent of powerful AI tools.

MISINFORMATION: PSYCHOLOGICAL INTERVENTIONS

If misinformation is considered a problem to be addressed in a given context, the question of solutions arises. Solutions need to be multi-pronged; from a policy perspective, there are at least four entry points for intervention:

- Regulatory (e.g., legislation, codes of conduct),
- Technological (e.g., algorithmic detection of problematic content on social-media platforms),
- Educational (e.g., systematic efforts to strengthen media and information literacy),
- Psychological (e.g., specific interventions targeting misinformation detection or sharing).

Our research has largely focused on the psychological dimension, where one of the significant issues we encounter is the resistance of misinformation to correction. This resistance stems from the inherent biases in human cognition and the difficulty and error-proneness of updating our memory and revising our existing knowledge, as correcting something that is believed to be true poses a cognitive challenge.

Accordingly, a substantial amount of research by our group and others has explored ways to effectively fact-check or debunk misinformation, which has highlighted important factors to consider. To illustrate, post-exposure corrections of misinformation are most effective when they incorporate the following elements:

1 Present a plausible alternative explanation. Learning that something isn't true can leave a gap in our understanding (e.g., if the vaccine did not cause the symptom, what did?); such gaps create psychological discomfort, so offering an alternative to fill the gap is important.

2 Explain in some detail why the misinformation is incorrect. There is no 'magic' number of counterarguments; all relevant arguments help.

3 Do not repeat the misinformation more than necessary, and warn recipients before exposing them to the to-be-corrected misinformation. While it is usually necessary and beneficial to repeat the misinformation once to ensure the correction is clear and salient, avoid additional repetition of the false claim, as familiarity tends to increase belief.

4 Expect the effectiveness of corrections to wane over time as memory for the correction fades. Be prepared to present a correction multiple times to enhance efficacy and longevity.

5 Achieve a credibility surplus. Ensure the correction comes from a source that is perceived as trustworthy by the recipient, and discredit the misinformation source where appropriate, e.g., by pointing out vested interests.

> **Practitioners must be aware that any intervention risks amplifying misinformation sources and 'buying into' their framing of an issue.**

These correction strategies should be incorporated into a larger intervention plan. Ideally, there should be ongoing monitoring of an information environment to enable an informed evaluation of the extent to which specific misinformation pieces are gaining traction and posing a risk of harm. Practitioners must be aware that any intervention risks amplifying misinformation sources and 'buying into' their framing of an issue. As such, debunking should only be applied after careful consideration of all potential outcomes.

ALTERNATIVE STRATEGIES

Since debunking can only ever operate retroactively, practitioners should consider alternative strategies. These include active promotion of truthful narratives and factual evidence, competence boosts, and behaviour-oriented nudges.

Competence boosts include educational tools to enhance media and information literacy skills, such as lateral reading, and inoculation interventions that aim to protect consumers from misinformation by explaining the misleading argumentation strategies that disinformants use in their persuasive attacks. Although further research is needed, one potential benefit of this approach is that inoculated individuals may be able to transfer the gained resilience to other topics. For example, understanding that a climate-change-denying argument uses cherry-picking tends to provide some protection against cherry-picked arguments in other domains, such as vaccination.

Behaviour nudges include accuracy prompts that remind the consumer to consider information veracity, the introduction of friction to reduce unwanted behaviour (i.e., sharing misinformation), and the use of social norms to highlight that most people try not to share misinformation and believe sharing misinformation is wrong.

To summarise, targeted corrections that follow our five recommendations can help counter (potentially) harmful misinformation where it arises and begins spreading. However, a whole array of evidence-based psychological strategies is available to practitioners, which cumulatively can contribute to a healthier information environment.

Ullrich Ecker is a Professor of Cognitive Psychology and Australian Research Council Future Fellow; Toby Prike is a Postdoctoral Research Associate; Li Qian Tay is a PhD Student; all authors are at the University of Western Australia's School of Psychological Science.

SARAH MARSDEN & JAMES LEWIS

HOW DO CASE MANAGEMENT TOOLS WORK TO COUNTER RADICALISATION?

When examining CVE interventions, people often ask "what works". Few have focussed on *how* they work. Here, Sarah Marsden and James Lewis present the latest research from their systematic review.

INTRODUCTION

One of the most commonly asked questions in the context of programmes to counter radicalisation or CVE interventions is what works to reduce the risk of radicalisation. Few have focussed their attention on understanding how interventions work. Rather than just assessing whether specific interventions such as ideological support or mentoring are effective, we were more concerned with understanding whether it matters how those interventions are delivered. To do that we searched through nearly 70,000 papers published on case management interventions to counter radicalisation to violence, in seven languages, to understand:

1. Whether the tools and approaches that are used to counter radicalisation to violence worked;

2. Whether they are implemented as they are intended to be; and

3. What factors influence how case management tools and approaches are implemented.

Case management: interventions that offer packages of support tailored to the specific needs of each individual from identification of a potential client through to their exit from a programme.

Tools: methods used to support the case management process such as case conferences or risk assessment processes.

Approaches: intervention logics or theories of change that underpin implementation and delivery. For example, the idea that interventions should be matched to someone's level of risk and be responsive to their needs.
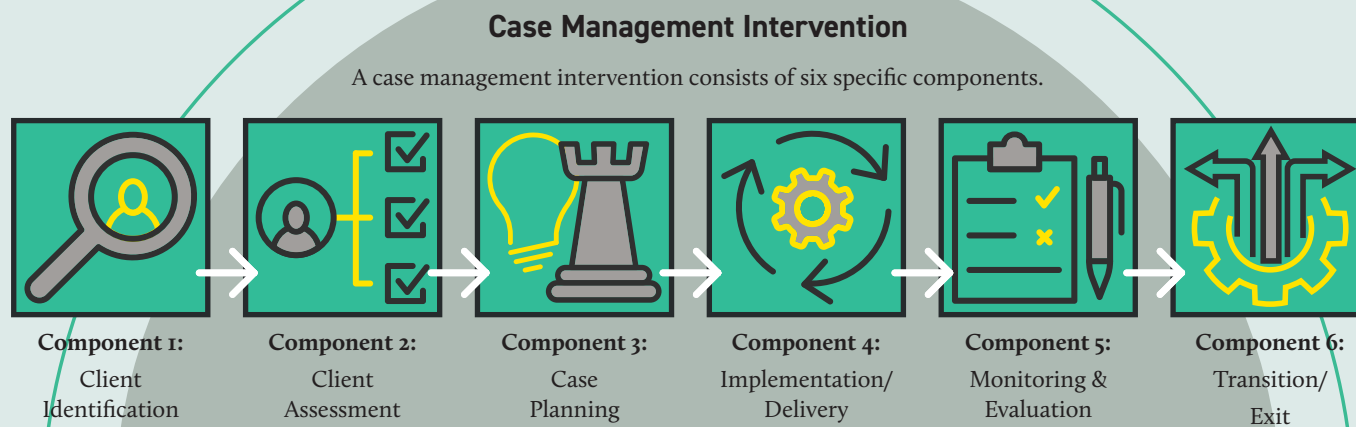
Case Management Intervention

A case management intervention consists of six specific components.

| Component 1: Client Identification | Component 2: Client Assessment | Component 3: Case Planning | Component 4: Implementation/ Delivery | Component 5: Monitoring & Evaluation | Component 6: Transition/ Exit |

Figure 1. The intensive case management process (based on NCMN, 2009)

FINDINGS

We didn't find any eligible studies that examined the effectiveness of case management interventions. However, the evidence base relating to implementation is more robust: 46 eligible studies examined the implementation of case management tools or approaches. These covered a range of different tools (see Table).

Stage	Tools and Methods Examined in Included Studies
1. Client Identification	• Outreach work post identification/referral
2. Client Assessment	• Client assessment tools • Multi-agency working • Case conferences
3. Case Planning	• Client assessment and case planning tools • Multi-agency working • Case conferences
4. Implementation / Delivery	• Tailoring intervention services and goals • Practitioner characteristics and approaches • Practitioner supervision and quality assurance
5. Monitoring & Evaluation	• Client assessment tools • Case file and case note data • Case conferences • Less structured qualitative data
6. Transition/ Exit	• Interagency coordination

Our analysis identified a number of factors that facilitated the implementation of case management processes. Efficient and effective multi-agency working, supported by strong and transparent relationships between partners was identified as a key facilitator of implementation. So too was practitioner experience and expertise: several studies highlighted how interventions benefited from being able to draw on relevant, interdisciplinary, case management and subject matter expertise.

We also identified a number of potential implementation barriers, most notably public and political factors, and resourcing constraints. The public and political scrutiny placed on counter-

> **Efficient and effective multi-agency working, supported by strong and transparent relationships between partners was identified as a key facilitator of implementation.**

radicalisation work can place pressure on practitioners, who operate in specific legislative contexts that influence how they conduct their work. Practitioners may also face economic and time constraints, particularly when interventions are financed through short-term funding.
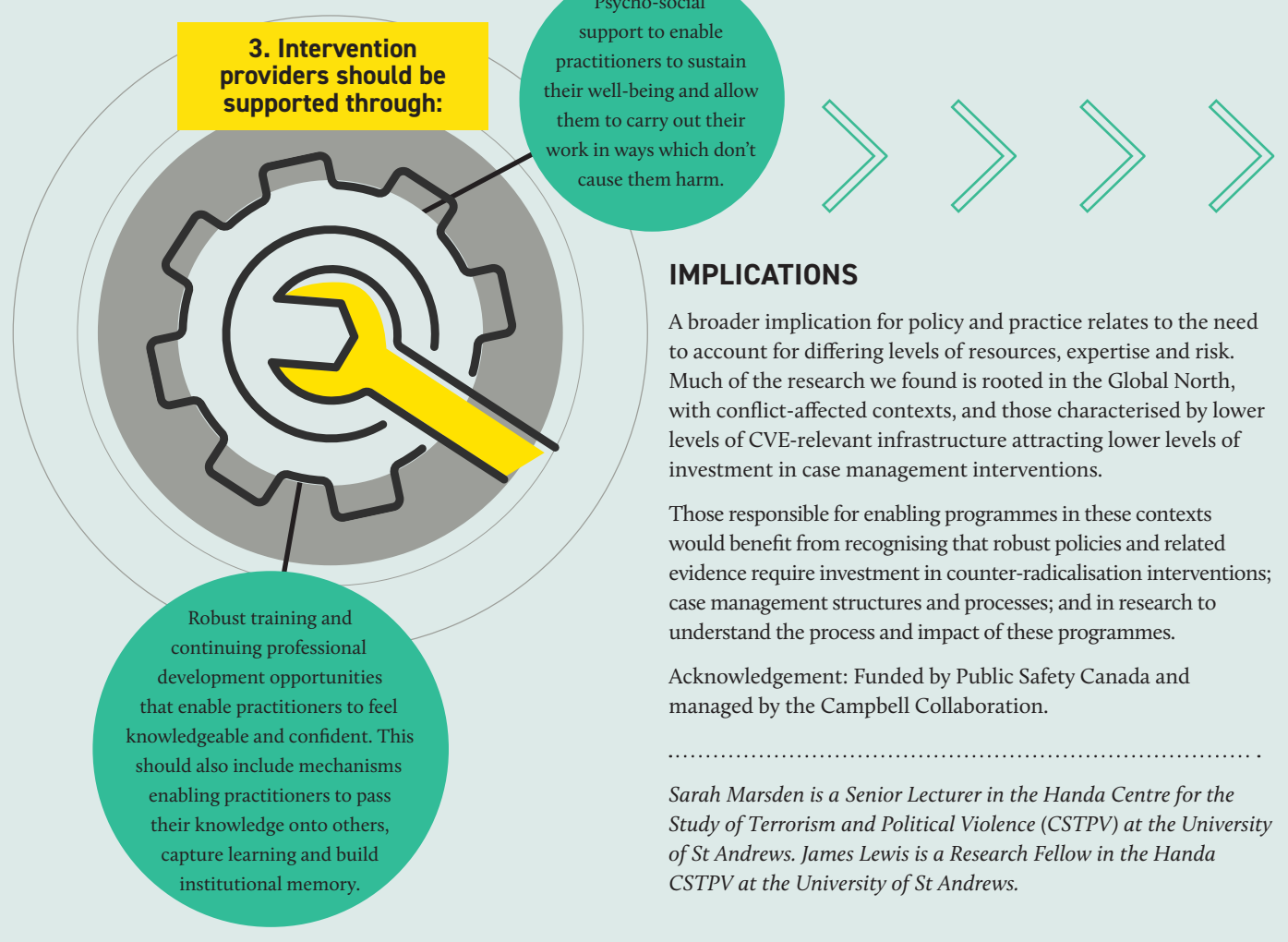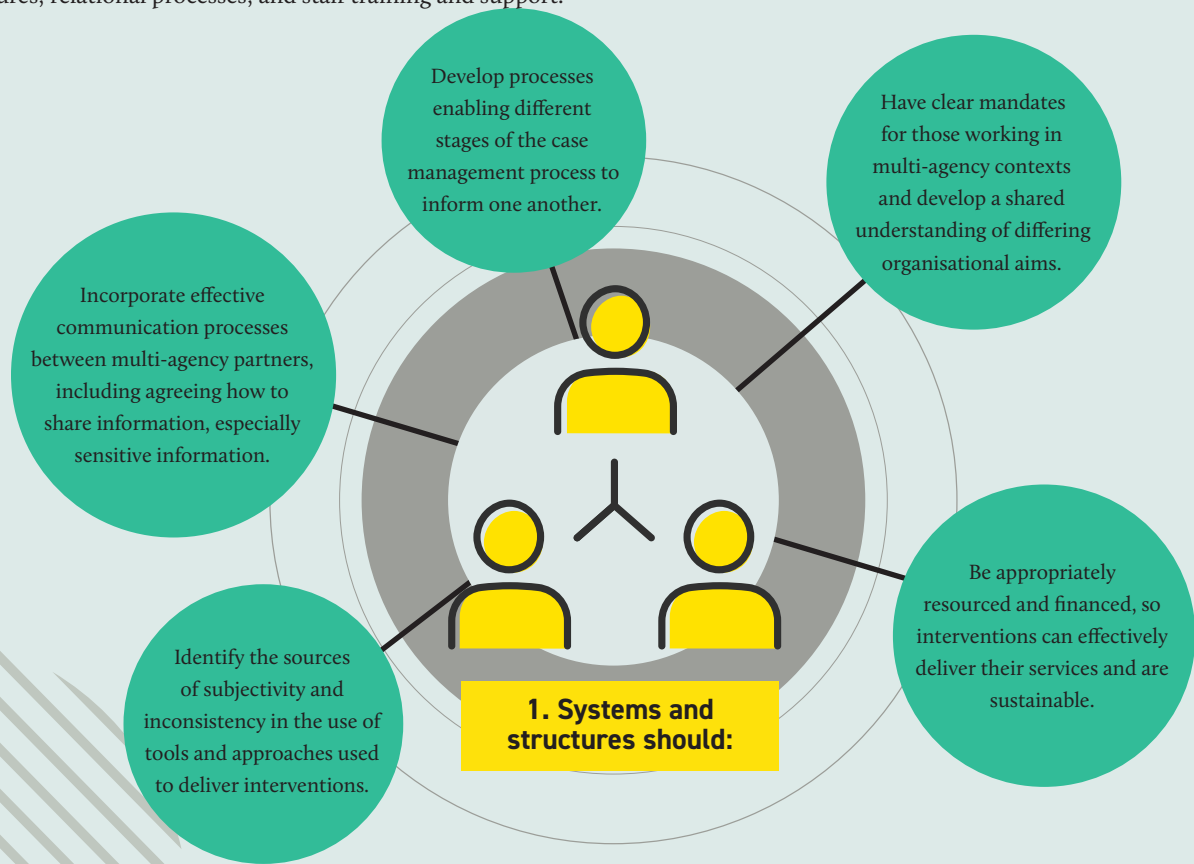
The research identified a number of factors that can shape how interventions are delivered in different contexts. Examples of these include whether an intervention is voluntary or mandated; the specific regional or national context; and the features of the settings in which the intervention is delivered, for instance whether it operates in a correctional or community context.

CONCLUSION

There is insufficient evidence to say whether the case management tools and approaches currently in use in the UK and elsewhere are effective. This points to the need for intervention policy to ensure monitoring and evaluation processes are built into programme design.

Nevertheless, there is a growing body of evidence highlighting those factors that can facilitate, or create barriers, to the delivery of counter-radicalisation interventions. This research is not yet robust. However, it points to three clusters of factors that offer insights into emerging good practice: the role of systems and structures; relational processes; and staff training and support.

> **There is a growing body of evidence highlighting those factors that can facilitate, or create barriers, to the delivery of counter-radicalisation interventions.**

1. Systems and structures should:

Develop processes enabling different stages of the case management process to inform one another.

Have clear mandates for those working in multi-agency contexts and develop a shared understanding of differing organisational aims.

Incorporate effective communication processes between multi-agency partners, including agreeing how to share information, especially sensitive information.

Be appropriately resourced and financed, so interventions can effectively deliver their services and are sustainable.

Identify the sources of subjectivity and inconsistency in the use of tools and approaches used to deliver interventions.

2. Relational processes should:

Develop ways of developing trust and building long-term relationships between multi-agency partners.

Address tensions that might emerge around differing organisational goals and priorities, e.g., in relation to rehabilitative and public protection related goals.

Identify the power hierarchies that are at work and collaborate with partners to acknowledge and address them.

Nurture and support relationships between clients and intervention providers or case workers.

3. Intervention providers should be supported through:

Psycho-social support to enable practitioners to sustain their well-being and allow them to carry out their work in ways which don't cause them harm.

Robust training and continuing professional development opportunities that enable practitioners to feel knowledgeable and confident. This should also include mechanisms enabling practitioners to pass their knowledge onto others, capture learning and build institutional memory.

IMPLICATIONS

A broader implication for policy and practice relates to the need to account for differing levels of resources, expertise and risk. Much of the research we found is rooted in the Global North, with conflict-affected contexts, and those characterised by lower levels of CVE-relevant infrastructure attracting lower levels of investment in case management interventions.

Those responsible for enabling programmes in these contexts would benefit from recognising that robust policies and related evidence require investment in counter-radicalisation interventions; case management structures and processes; and in research to understand the process and impact of these programmes.

....

Sarah Marsden is a Senior Lecturer in the Handa Centre for the Study of Terrorism and Political Violence (CSTPV) at the University of St Andrews. James Lewis is a Research Fellow in the Handa CSTPV at the University of St Andrews.

NIKKI POWER, RICHARD PHILPOT & MARK LEVINE

THE PSYCHOLOGY OF INTEROPERABILITY:
IMPROVING EMERGENCY SERVICES' TEAMWORK

When disaster strikes, different first responder groups must unite and coordinate their efforts. Recognising the psychological dynamics of interoperability is vital to support joint working and optimise life-saving.

The emergency services must operate as an efficient, interoperable team during crises. Effective joint working enables team members to combine their knowledge, skills and expertise in pursuit of collective goals; namely, saving lives. However, the nature of emergencies challenges teamwork. Teams have to operate under conditions of extreme uncertainty while managing multiple competing task demands. Imagine that an explosion has occurred in a busy city centre. The Police face the urgent task of identifying and neutralising any threats, while the Fire Service tackles fire outbreaks and assesses the risk of structural collapses in buildings. Simultaneously, the Ambulance Service must work to administer medical treatment and transport patients to hospitals. However, these tasks do not unfold in a linear fashion, as the success of one team's mission often depends on the performance of another. For instance, paramedics cannot administer treatment unless the Police have effectively neutralised further threats and the Fire Service has dealt with any fires. In this intricate web of interdependent tasks, a cohesive approach to interoperability is paramount to ensure the seamless coordination of activities in pursuit of the collective goal of saving lives.

THE PSYCHOLOGY OF INTEROPERABILITY

Despite the importance of interoperability, its integration within the emergency services has proven challenging. JESIP (the joint emergency services interoperability programme) was launched in the UK in 2012 with the goal of improving interoperability, but the recent public inquiry into the Manchester Arena attack concluded that interoperability has not been embedded. We argue that this failure to fully integrate interoperability is due to a poor understanding of psychology. Teams are social units, which makes them inherently psychological. So if interoperability is to be seamlessly woven within the social fabric of the emergency services then a robust understanding of psychology is essential. Developing this understanding is the goal of our CREST-funded research.

DEFINING INTEROPERABILITY

A first step in our research was defining interoperability. Existing definitions were often vague and abstract, and so we sought to build a new definition of interoperability that identified its precise components. To do this, we conducted a systematic review of the literature and identified five components that are essential for interoperability to thrive.

> "We define interoperability as a shared system of technology and teamwork built upon trust, identification, goals, communication, and flexibility.

1. Communication

Interoperable teams must be able to prioritise efficient and meaningful communication, sharing relevant information while avoiding overload. This fosters a shared understanding and informed decision-making.

2. Flexibility

Successful interoperable teams must embrace flexibility and decentralisation. Team members should have a clear understanding of roles, which will enable adaptivity and effective action, even in the absence or overload of another team member. Flexible and decentralised teams allow responsibilities and decision-making authority to be distributed, aiding an efficient response.

3. Trust

Trust is a foundational element within an interoperable team. It encompasses different dimensions, including interpersonal trust (based on personal familiarity), role-based trust (having confidence in the competence and reliability of individuals to fulfil a specific role), and group-level trust (which extends to any member representing a particular organisation or profession). Establishing and maintaining trust within the team is vital for interoperability.

4. Identity

Emergency workers must maintain secure organisational identities within their interoperable team. Organisational identity refers to how individuals perceive themselves as members of their organisation and their sense of alignment with its mission and values. Efforts to promote interoperability, through changes in doctrine or training, should prioritise respecting team members' identities to prevent identity threats and promote a positive, inclusive team environment.

5. Goals

Interoperable teams must have cohesive goals. While the overarching aim is saving lives, practical implementation may vary across roles and services. Translating and aligning goals enables harmonious teamwork and coordination, fostering a unified team.

Taken together, we define interoperability as a shared system of technology and teamwork built upon trust, identification, goals, communication, and flexibility. This definition can be used to empower researchers and end-users alike, pinpointing the vital components around which to design targeted interoperability research and training.

HOW DO WE TRAIN INTEROPERABILITY?

A second step in our research was to provide practical recommendations for how to train interoperability; identifying specific methods that can be used to, for example, build trust between team members. Through interviews with commanders from the Police, Fire and Ambulance Service, we found that regular face-to-face joint training is crucial for building knowledge, skills, breaking down barriers, fostering interpersonal and group-level trust, and promoting a shared sense of togetherness and collective teamwork. However, commanders acknowledged that resource and funding pressures make this type of training rare. Simulations were deemed a valuable compromise, offering time-efficient and resource-friendly opportunities for collaboration and exercising. Yet, even planned training sessions are frequently disrupted as Police and Paramedics are pulled away last minute to cope with operational demands.

SO WHAT'S NEXT?

To enhance interoperability, it is essential to develop training programs that are explicitly tailored to foster social psychological connectiveness. The result of which will be stronger, more interconnected, efficient and robust emergency services. Yet, the realisation of this vision relies upon adequate investment in the Emergency Services; without such investment, this transformative goal will remain elusive.

Dr Nikki Power is a senior lecturer in organisational behaviour at the University of Liverpool. Dr Richard Philpot is a lecturer of Psychology at Lancaster University. Professor Mark Levine is a professor of Social Psychology at Lancaster University.

For more outputs (including an animation, guide, and poster) from this CREST-funded project visit **crestresearch.ac.uk/interop**

Image credit: © GWatkins | stock.adobe.com

MUHSIN YESILADA & PAUL GRASBY

A-Z of
MISINFORMATION

Do you know your algorithms from your Gen-Z? This A to Z provides examples of how false or misleading information can be spread and ways to combat it.

ALGORITHMS

Social media algorithms can amplify the spread of misinformation via recommender systems. The algorithms are curated to provide recommendations per the users' interests, search history and likes. So, users may be exposed to further misinformation if the algorithm identifies them engaging with the content.

BOTS

Misinformation can be disseminated by humans and automated online accounts, known as bots. Bots are widespread on social media platforms and, by emulating human social interactions, can mimic human users on platforms such as X (formerly Twitter), engage in excessive posting, retweeting of polarising news content, and reference influential figures. Not all bots are 'bad', but they can be used to amplify misinformation to manipulate political discourse.
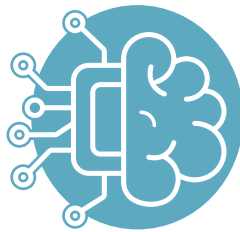
COGNITIVE BIAS

The underlying mechanism that leads to accepting misinformation. Well-known cognitive biases include familiarity (relying on information that is already familiar), availability (relying on information that comes readily to mind), and confirmation (processing information in a way that supports previous beliefs) which can influence an individual's judgements and decision-making when differentiating between factual information and fallacious claims.

DISINFORMATION

False information that is spread intentionally. Disinformation is often used to distort public perception for personal or political gain. Extremist groups often use false information to garner public support, invoking fear and recruiting members through manipulation.

EXPOSURE

Exposure and sharing are connected, yet they are distinct concepts. Most people share a small percentage of the material they are exposed to. Therefore, looking at what someone shares gives a restricted view of their information environment.

FAKE NEWS

Often used to describe false information. However, the term is less descriptive and useful than other terms such as 'disinformation' or 'misinformation'. Fake news has also ironically been used to politicise and justify misinformation, where the term is used to denigrate factually correct information produced by opponents.

GAMES

Prebunking describes proactively refuting false information before people fall for it. One way in which this can be done is through educational games. These games teach people about possible manipulation techniques that might be used against them.

HOSTILE STATE ACTORS

Hostile states spread misinformation by investing in public discourse influence. These nations may believe they will benefit from such initiatives. One example is the well-evidenced activities of the Russian, St. Petersburg based Internet Research Agency.

INOCULATION

If we conceptualise misinformation as akin to a virus that spreads through society, we can inoculate against it. Psychological inoculations aim to give people the 'mental antibodies' to resist persuasion from misinformation. Inoculation works by pre-emptively warning people that they might be manipulated and then giving them the skills to identify misinformation.

JOURNALIST INVESTIGATIONS

Individuals with journalistic backgrounds do much of the work on fact-checking and debunking. Examples include BBC Reality Check, Bellingcat and PolitiFact.com, operated by the Poynter Institute.

KNOWLEDGE REVISION

Even after receiving a correction and accepting it as true, misinformation can continue to shape people's beliefs. The continued influence effect (CIE) describes this process. Beyond the laboratory, CIE has been demonstrated for real-world events such as the 2003 Iraq war WMD, and vaccines and autism.

LIKES

A form of engagement on social media where the user shows others that they like the content posted by simply clicking a button. The visible likes and shares counter can significantly influence interaction with low-credibility information. People are more likely to share questionable content and less likely to fact-check it as engagement rises.

MISINFORMATION

The sharing of inaccurate and misleading information unintentionally. The rise of social media has brought concerns over misinformation to the fore, with the number of academic and policy-related articles on misinformation showing an exponential increase over time. As with any messaging, when considering misinformation, the source, message, context, and receiver are all important considerations in analysing its likely impact.

NEW MEDIA LITERACY

A major component in combating misinformation over the long term is new media literacy strategies. This educational intervention aims to improve people's ability to discern accurate and inaccurate news content on social media. For example, media literacy interventions might teach people to identify low-credibility news sources.
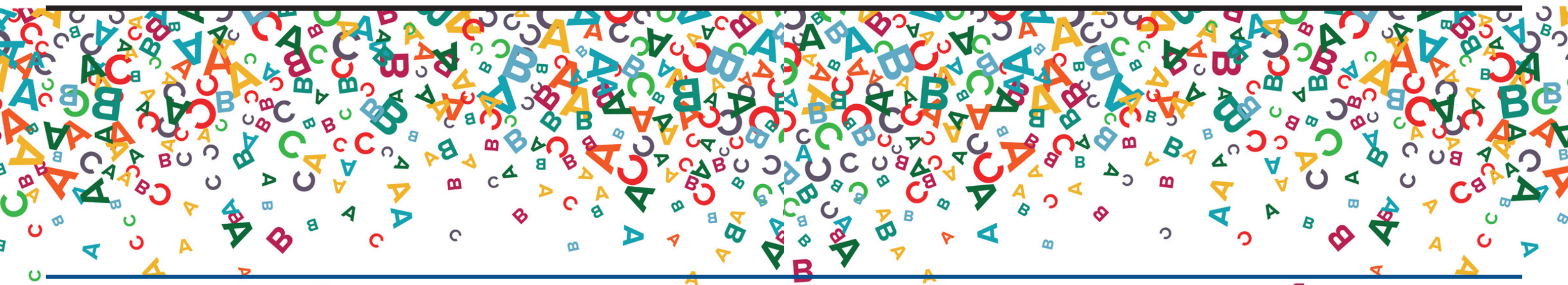
ONLINE SAFETY BILL

Final amendments were made to the bill during the third reading in the House of Lords on Wednesday 6th September 2023. The bill now returns to the House of Commons for further consideration. The Online Safety Bill is to quote a 'new set of laws to protect children and adults online. It will make social media companies more responsible for their users' safety on their platforms - HMG'. It would mandate that search engines and "user-to-user" applications filter illegal content. How the bill will effectively tackle mis - and dis - information is the subject of considerable debate.

PERSUASION

Persuasion can be used to disseminate misinformation by skillfully and convincingly presenting false or misleading information. Misinformation can be spread through persuasive techniques that appeal to people's emotions, biases, and preconceived beliefs, making them more susceptible to accepting inaccurate or misleading claims. Understanding the principles of persuasion is crucial for recognizing and countering misinformation

QANON

QAnon, an American political conspiracy theory and political movement, which had its first supporting congresswoman in Georgia, Republican Marjorie Taylor Greene. In a recent study by Wu and colleagues, official Republican or Democratic condemnation of Greene decreased positive views of QAnon but not Greene. The authors conclude that their "results suggest that public officials have a unique responsibility to criticize misinformation, but they also highlight the difficulty in shifting attitudes toward politicians who embrace and spread falsehoods" (whether intentional or unintentional).

RUMOURS

Rumours are often the breeding ground for political misinformation and conspiracy theories. Rumours can spread misinformation, leading individuals to believe and share false information. Peterson has drawn attention to the power of 'hostile political rumours', which can shape political outcomes by inciting hostility toward a specific politician or political group even when factual evidence for the rumour is scant.

SPOOFING

Where someone falsely claims to be someone else or falsely adopts a social standing or identity. Information spoofing includes falsifying, suppressing, or amplifying messages and may serve to influence public understanding of events. Spoofing (together with 'truthing' and 'social proofing') on digital platforms was observed by Innes and colleagues following the 2017 UK terrorist attacks

TRUST

Trust in the context of misinformation relates to the complex relationship between social media platforms, policy-makers, and users. Users generally express trust concerns regarding misinformation and data use, censorship, freedom of speech, and the interplay between these issues.

USERS

Social media users and their behaviour are integral when analysing the spread of misinformation. In one large, well-known longitudinal study, Vosoughi and colleagues found that human users (and not bots) were responsible for the dramatic spread of false news online, which was 70% more likely to be retweeted. The authors suggested that the degree of novelty and the emotional reactions of recipients may be responsible for the rapid and widespread online diffusion of falsehoods over the truth.

VISUAL DISINFORMATION, MEMES AND DEEP FAKES

Weikmann and Lecheler have noted that visual disinformation may be classed along two dimensions: (1) audio-visual richness, i.e., whether static or moving pictures are employed, and (2) manipulation sophistication, i.e., whether low-level or high-level creative techniques are used. Memes can be an important vehicle for spreading misinformation and will often aim to invoke emotions such as fear, anger and empathy by using humour. Recent times have seen the emergence of deep fakes utilising machine learning, making discerning true and false audio-visual information difficult.

WORLDVIEWS

Worldviews influence misinformation spread and reception. People will likely believe and disseminate misinformation that matches their values and views. These established worldviews can encourage false information and hinder the critical examination of data.

XENOPHOBIA

Misinformation can promote xenophobic attitudes. Xenophobic misinformation is particularly prominent in false narratives regarding migration and refugees. Xenophobic misinformation narratives in this context falsely claim that refugees and migrants are a danger to society.

YOUTUBE

As a social media platform with billions of daily views, YouTube has the potential to aid and worsen the spread of misinformation. Several investigations have examined whether the YouTube recommender system facilitates pathways to misinformation content. However, due to the limitations of these studies, such as algorithmic access, it is difficult to conclusively analyse this issue.

Z GENERATION Z (18-24 YRS)

The Reuters Institute at Oxford University has suggested that social networks have steadily replaced news websites as a primary source for younger audiences overall, with Instagram, TikTok, and YouTube becoming increasingly popular for news amongst this group. In 2022, 39% of 18–24s used social media as their main news source, compared with 34% who preferred to go directly to a news website or app. Younger audiences were also the lowest-trusting age groups, with only a third 'trusting most news most of the time', with substantial rises in avoidance of the news compared with older age groups. Whether Gen Z, with different internet habits, will be more or less prone to disinformation remains a considerable debate.

Muhsin Yesilada is a Doctor of Philosophy in the School of Psychological Science at the University of Bristol. Paul Grasby is a Research to Practice Fellow at CREST. The authors wish to thank Professor Tom Buchanan and Professor Steven Lewandowsky for their initial advice on the terms to include in this piece.

READ MORE

Read more about some of the research that our contributors mention in their articles. We've flagged up those that are open access and given links to online versions where they are available. For full references and citations please visit the online version at *crestresearch.ac.uk/magazine/misinformation*

TOM BUCHANAN, ROTEM PERACH & DEBORAH HUSBANDS: WHY DO PEOPLE SHARE FALSE POLITICAL INFORMATION ONLINE?

Buchanan, T. (2020) Why Do People Share Disinformation On Social Media? *https://crestresearch.ac.uk/resources/disinformation-on-social-media/*

Buchanan, T. (2020) Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS One,* 15(10). *https://doi.org/10.1371/journal.pone.0239666*

Buchanan, T., & Kempley, J. (2021) Individual differences in sharing false political information on social media: Direct and indirect effects of cognitive-perceptual schizotypy and psychopathy. *Personality and Individual Differences,* 182. *https://doi.org/10.1016/j.paid.2021.111071*

Copeland, S. & Marsden, S. (2020) The Relationship Between Mental Health Problems and Terrorism. *https://crestresearch.ac.uk/resources/the-relationship-between-mental-health-problems-and-terrorism/*

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022) The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology,* 1, 13-29. *https://www.nature.com/articles/s44159-021-00006-y*

Gill, P., Corner, E., McKee, A., Hitchen, P., & Betley, P. (2022) What Do Closed Source Data Tell Us About Lone Actor Terrorist Behavior? A Research Note. *Terrorism and Political Violence,* 34(1), 113-130. *https://doi.org/10.1080/09546553.2019.1668781*

Lewandowsky, S., Van der Linden, S., & Cook, J. (2018) Can We Inoculate Against Fake News? *https://crestresearch.ac.uk/comment/can-we-inoculate-against-fake-news/*

Modirrousta-Galian, A. & Higham, P. A. (2023) Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology General,* 152(9), 2411–2437. *https://doi.org/10.1037/xge0001395*

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature,* 592, 590–595. *https://doi.org/10.1038/s41586-021-03344-2*

Perach, R., Joyner, L., Husbands, D., & Buchanan, T. (2023) Why Do People Share Political Information and Misinformation Online? Developing a Bottom-Up Descriptive Framework. *Social Media + Society,* 9(3). *https://doi.org/10.1177/20563051231192032*

WILLIAM DANCE: ADDRESSING ALGORITHMS IN DISINFORMATION

Baker, P., Gabrielatos, C., & McEnery, T. (2013) *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press.* Cambridge: Cambridge University Press. *https://doi.org/10.1017/CBO9780511920103*

Coltman-Patel, T., Dance, W., Demjén, Z., Gatherer, D., Hardaker, C., & Semino, E. (2022) 'Am I being unreasonable to vaccinate my kids against my ex's wishes?' – A corpus linguistic exploration of conflict in vaccination discussions on Mumsnet Talk's AIBU forum. *Discourse, Context & Media,* 48. *https://doi.org/10.1016/j.dcm.2022.100624*

Hardaker, C. & McGlashan, M. (2016) "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics,* 91, 80-93. *https://doi.org/10.1016/j.pragma.2015.11.005*

Jones-Jang, S. M., Kim, D. H., & Kenski, K. (2020) Perceptions of mis- or disinformation exposure predict political cynicism: Evidence from a two-wave survey during the 2018 US midterm elections. *New Media & Society,* 23(10), 3105-3125. *https://doi.org/10.1177/1461444820943878*

Vaccari, C. & Chadwick, A. (2020) Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society,* 6(1). *https://doi.org/10.1177/2056305120903408*

ULLRICH ECKER, TOBY PRIKE & LI QIAN TAY: PSYCHOLOGICAL INTERVENTIONS TO COMBAT MISINFORMATION

McCright, A.M., Dunlap, R.E. (2017) Combatting misinformation requires recognising its types and the factors that facilitate its spread and resonance. *Journal of Applied Research in Memory and Cognition,* 6(4), 389-396. *https://doi.org/10.1016/j.jarmac.2017.09.005*

Lewandowsky, S., Ecker, U., & Cook, J. (2017) Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition,* 6(4), 353–369. *https://doi.org/10.1016/j.jarmac.2017.07.008*

Loomba, S., de Figueiredo, A., Piatek, S.J., de Graaf, K., & Larson, H.J. (2021) Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the U.K. and USA. *Nature Human Behaviour,* 5, 337–348. *https://doi.org/10.1038/s41562-021-01056-1*

Bursztyn, L., Rao, A., Roth, C.P., & Yanagizawa-Drott, D.H. (2020) Misinformation during a pandemic. *National Bureau of Economic Research.* *https://www.nber.org/papers/w27417*

Simonov, A., Sacher, S., Dubé, J.P., & Biswas S. (2022) Frontiers: The Persuasive Effect of Fox News: Noncompliance with Social Distancing During the COVID-19 Pandemic. *Marketing Science,* 41(2), 230-242. *http://dx.doi.org/10.2139/ssrn.3600088*

Buchanan, B., Lohn, A., Musser, M., & Sedova, K. (2021) *Truth, lies, and automation: How language models could change disinformation.* Center for Security and Emerging Technology. *https://doi.org/10.51593/2021CA003*

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020) Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest,* 21(3):103–56. *https://doi.org/10.1177/1529100620946707*

Lewandowsky, S., Cook, J., Ecker, U., Albarracin, D., Amazeen, M., Kendou, P., et al (2020) *The Debunking Handbook 2020.* *https://doi.org/10.17910/b7.1182*

Lewandowsky, S. & van der Linden, S. (2021) Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology,* 32(2), 348-384. *https://doi.org/10.1080/10463283.2021.1876983*

RICKY GREEN, IMANE KHAOUJA, DANIEL TORIBIO-FLÓREZ & KAREN M. DOUGLAS: CONSPIRACY THEORIES: THEIR PROPAGATION AND LINKS TO POLITICAL VIOLENCE

Douglas, K.M., Sutton, R.M., & Cichocka, A. (2017) The psychology of conspiracy theories. *Current Directions in Psychological Science,* 26(6), 538-542 *https://doi.org/10.1177/0963721417718261*

Imhoff, R., Zimmer, F., Klein, O., António, J. H., Babinska, M., Bangerter, A., Bilewicz, M., Blanuša, N., Bovan, K., Bužarovska, R., Cichocka, A., Delouvée, S., Douglas, K. M., Dyrendal, A., Etienne, T., Gjoneska, B., Graf, S., Gualda, E., Hirschberger, G., ... Van Prooijen, J. (2022) Conspiracy mentality and political orientation across 26 countries. *Nature Human Behaviour,* 6(3), 392-403. *https://doi.org/10.1038/s41562-021-01258-7*

Kruglanski, A. W., Molinario, E., Ellenberg, M., & Di Cicco, G. (2022) Terrorism and conspiracy theories: A view from the 3N model of radicalization. *Current Opinion in Psychology,* 47, 101396. *https://doi.org/10.1016/j.copsyc.2022.101396*

Marcellino, W., Helmus, T. C., Kerrigan, J., Reininger, H., & Karimov, R. I. (2021) *Detecting conspiracy theories on social media: Improving machine learning to detect and understand online conspiracy theories.* *https://www.rand.org/pubs/research_reports/RRA676-1.html*

Rottweiler, B. & Gill, P. (2020) Conspiracy beliefs and violent extremist intentions: The contingent effects of self-efficacy, self-control and law-related morality. *Terrorism and Political Violence,* 34(7), 1485-1504. *https://doi.org/10.1080/09546553.2020.1803288*

Rousis, G. J., Richard, F. D., & Wang, D. D. (2020) The truth is out there: The prevalence of conspiracy theory use by radical violent extremist organizations. *Terrorism and Political Violence,* 34(8), 1739-1757. *https://doi.org/10.1080/09546553.2020.1835654*

Sutton, R. M. & Douglas, K. M. (2022) Rabbit hole syndrome: Inadvertent, accelerating, and entrenched commitment to conspiracy beliefs. *Current Opinion in Psychology,* 48, 101462. *https://doi.org/10.1016/j.copsyc.2022.101462*

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014) Analytic thinking reduces belief in conspiracy theories. *Cognition,* 133(3), 572-585. *https://doi.org/10.1016/j.cognition.2014.08.006*

University of Cambridge, BBC Media Action, & Jigsaw (2022) *A Practical Guide to Prebunking Misinformation.* Info interventions. *https://interventions.withgoogle.com/static/pdf/A_Practical_Guide_to_Prebunking_Misinformation.pdf*

van der Linden, S. & Roozenbeek, J. (2020) Psychological inoculation against fake news. In Greifeneder, R., Jaffe, M., Newman, E., & Schwarz, N. (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 147-169). Routledge. *http://dx.doi.org/10.4324/9780429295379*

HELEN INNES, ANDREW DAWSON & MARTIN INNES: OSINT VS DISINFORMATION: THE INFORMATION THREATS 'ARMS RACE'

Innes, M., Dobreva, D. & Innes, H. (2021) Disinformation and digital influencing after terrorism: spoofing, truthing and social proofing. *Contemporary Social Science,* 16 (2), pp. 241-255. *https://doi.org/10.1080/21582041.2019.1569714*

Louart, M., Szkolnik J., Boudraa, A., Le Lann, J., & Le Roy, F. (2023) Detection of AIS Messages Falsifications and Spoofing by Checking Messages Compliance with TDMA Protocol. *Digital Signal Processing,* 136. *https://doi.org/10.1016/j.dsp.2023.103983*

US Maritime Advisory (2022) *GPS Interference & AIS Spoofing.* *https://www.maritime.dot.gov/msci/2022-005-various-gps-interference-ais-spoofing*

Zegart, A. (2022) Open Secrets: Ukraine and the Next Intelligence Revolution. Foreign Affairs. *https://www.foreignaffairs.com/world/open-secrets-ukraine-intelligence-revolution-amy-zegart*

GRACE MCKENZIE: HIDING IN PLAIN SITE

Bond, C. F., Jr., & DePaulo, B. M. (2006) Accuracy of deception judgments. *Personality and Social Psychology Review,* 10, 214-234. *https://doi.org/10.1207/s15327957pspr1003_*

Chavoshi, N., Hamooni, H., & Mueen, A. (2016) Identifying correlated bots in Twitter. In: Spiro E., Ahn YY. (eds) *Social Informatics.* SocInfo 2016: Lecture Notes in Computer Science, vol 10047. Springer, Cham. *https://doi.org/10.1007/978-3-319-47874-6_2*

Kudugunta, S. & Ferrara, E. (2018) Deep neural networks for bot detection. *Information Sciences,* 467, 312-322. *https://arxiv.org/pdf/1802.04289.pdf*

Meta (2022, May 31) *Fake Accounts, Community Standards Enforcement Report.* https://transparency.fb.com/data/community-standards-enforcement/fake-accounts/facebook/#content-actioned

Statista (2023, May 30) *Number of internet and social media users worldwide as of April 2023.* https://www.statista.com/statistics/617136/digital-population-worldwide/

Statista (2023, May 30) *Most popular social networks worldwide as of January 2023.* https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

SARAH MARSDEN & JAMES LEWIS: HOW DO CASE MANAGEMENT TOOLS WORK TO COUNTER RADICALISATION?

Cherney, A., & Belton, E. (2021) Evaluating case-managed approaches to counter radicalization and violent extremism: An example of the Proactive Integrated Support Model (PRISM) intervention. *Studies in Conflict & Terrorism,* 44(8), 625-645. https://doi.org/10.1080/1057610X.2019.1577016

Lewis, J., Marsden, S., Cherney, A., Zeuthen, M., Bélanger, J. J., Zubareva, A., Brandsch, J., & Lubrano, M. (2023) PROTOCOL: Case management interventions seeking to counter radicalisation to violence: A systematic review of tools and approaches. *Campbell Systematic Reviews,* 19, https://doi.org/10.1002/cl2.1301

ISABELLA ORPEN: USING FUZZY SET QUALITATIVE COMPARATIVE ANALYSIS TO EXAMINE HETEROGENEITY IN CONSPIRACY BELIEVERS

Armaly, M. T., Buckley, D. T., & Enders, A. M. (2022) Christian nationalism and political violence: Victimhood, racial identity, conspiracy, and support for the capitol attacks. *Political Behavior,* 44(2), 937-960. https://doi.org/10.1007/s11109-021-09758-y

Bartlett, J. & Miller, C. (2010) *The power of unreason: Conspiracy theories, extremism and counter-terrorism.* London: Demos. https://demos.co.uk/research/the-power-of-unreason/

Daniel, L. & Harper, D. J. (2022) The social construction of conspiracy beliefs: A Q-methodology study of how ordinary people define them and judge their plausibility. *Journal of Constructivist Psychology,* 35(2), 564-585. https://doi.org/10.1080/10720537.2020.1837695

Lee, S. S. (2014) Using fuzzy-set qualitative comparative analysis. *Epidemiology and health,* 36. https://doi.org/10.4178/epih/e2014038

Pigden, C. (2006) "Complots of Mischief." In D. Coady (ed.), *Conspiracy Theories: The Philosophical Debate,* pp. 139–66. Aldershot: Ashgate. https://doi.org/10.1177/0048393122108100I

van Prooijen, J. W. & Douglas, K. M. (2017) Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies,* 10(3), 323–333. https://doi.org/10.1177/1750698017701161

Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012) Dead and alive: Beliefs in contradictory conspiracy theories. *Social psychological and personality science,* 3(6), 767-773. https://doi.org/10.1177/1948550611434786

DIMITRI PAVLOUNIS & KELSEY DAVIS: WHAT KIND OF DIGITAL MEDIA LITERACY?: BUILDING STUDENT RESILIENCE TO MISINFORMATION THROUGH EVIDENCE-BASED APPROACHES

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021) Students' Civic Online Reasoning: A national portrait. *Educational Researcher.* https://purl.stanford.edu/cz440cm8408

Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P., Batson, M., Grosso, R., Matthews, M., Miller, V., & Caulfield, M. (2021) Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course. Cognitive Research: *Principles and Implications,* 6(1). https://doi.org/10.1186/s41235-021-00291-4

Caulfield, M. (2018) A short history of CRAAP. HAPGOOD. https://hapgood.us/2018/09/14/a-short-history-of-craap/

Kohnen, A. M., Mertens, G. E., & Boehm, S. M. (2020). Can middle schoolers learn to read the web like experts? Possibilities and limits of a strategy-based intervention. *Journal of Media Literacy Education,* 12(2), 64-79. https://doi.org/10.23860/JMLE-2020-12-2-6

Pavlounis, D., Johnston, J., Brodsky, J., & Brooks, P. (2021) *The Digital Media Literacy Gap: How to build widespread resilience to false and misleading information using evidence-based classroom tools.* CIVIX Canada. https://ctrl-f.ca/en/the-evidence/

Wineburg, S., Breakstone, J., Ziv, N., & Smith, M. (2020) E*ducating for Misunderstanding: How Approaches to Teaching Digital Literacy Make Students Susceptible to Scammers, Rogues, Bad Actors, and Hate Mongers* (Working Paper A-21322). Stanford History Education Group. Stanford University. https://purl.stanford.edu/mf412bt5333

Wineburg, S., & McGrew, S. (2019) Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information. *Teachers College Record,* 121(11), 1–40. https://doi.org/10.1177/016146811912101102

NIKKI POWER, RICHARD PHILPOT & MARK LEVINE: THE PSYCHOLOGY OF INTEROPERABILITY: IMPROVING EMERGENCY SERVICES' TEAMWORK

Power, N., Alcock, J., Levine, M., & Philpot, R. (2023) The Psychology of Interoperability: Study One Summary. CREST Guide. https://crestresearch.ac.uk/resources/the-psychology-of-interoperability-study-one/

Saunders, J. (2022) Manchester Arena Inquiry. Volume 2: Emergency Response. https://manchesterarenainquiry.org.uk/report-volume-two/

ZOEY REEVE: MISLEADING A GROUP TO INEFFECTIVENESS

Felps, W., Mitchell, T, R., & Byington, E. (2006) How, When, and Why Bad Apples Spoil the Barrel: Negative Group Members and Dysfunctional Groups. *Research in Organisational Behaviour,* 27: 175-222. https://doi.org/10.1016/S0191-3085(06)27005-9

Gonzales, A. L., Hancock, J. T., & Pennebaker, J, W. (2010) Language Style Matching as a Predictor of Social Dynamics in Small Groups. *Communication Research,* 37(1): 3-19. https://doi.org/10.1177/0093650209351468

Ormerod, T., Barrett, E., & Taylor, P. (2012) Investigative sense-making in criminal contexts. *Proceedings of the Seventh International NDM Conference* (Ed. J.M.C Schraagen), Amsterdam, The Netherlands. https://ris.utwente.nl/ws/portalfiles/portal/211824698/Investigative_sense_making_in_criminal_contexts.pdf

BETTINA ROTTWEILER & PAUL GILL: CONSPIRATORIAL THINKING AND FAR-RIGHT EXTREMIST ATTITUDES

Hebel-Sela, S., Hameiri, B., & Halperin, E. (2022) The vicious cycle of violent intergroup conflicts and conspiracy theories. *Current Opinion in Psychology,* 47, 101422. https://doi.org/10.1016/j.copsyc.2022.101422

Jolley, D., Marques, M. D., & Cookson, D. (2022) Shining a spotlight on the dangerous consequences of conspiracy theories. *Current Opinion in Psychology,* 47, 101363. https://doi.org/10.1016/j.copsyc.2022.101363

Obaidi, M., Kunst, J., Ozer, S., & Kimel, S. Y. (2022) The "Great Replacement" conspiracy: How the perceived ousting of Whites can evoke violent extremism and Islamophobia. *Group Processes & Intergroup Relations,* 25(7), 1675-1695. https://doi.org/10.1177/01461672231167694

Rottweiler, B. & Gill, P. (2022) Conspiracy beliefs and violent extremist intentions: The contingent effects of self-efficacy, self-control and law-related morality. *Terrorism and Political Violence,* 34(7), 1485-1504. https://doi.org/10.1080/09546553.2020.1803288

SANDER VAN DER LINDEN: A PSYCHOLOGICAL VACCINE AGAINST MISINFORMATION

van der Linden, S. (2023) *Foolproof: Why We Fall for Misinformation and How to Build Immunity,* W.W. Norton & Company Ltd. Inc. https://www.sandervanderlinden.com/

MUHSIN YESILADA & PAUL GRASBY: A-Z OF MISINFORMATION

Basol, M., Roozenbeek, J., & Van der Linden, S. (2020) Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition,* 3(1). https://doi.org/10.5334/joc.91

Innes M., Dobreva, D. & Innes, H. (2021) Disinformation and digital influencing after terrorism: spoofing, truthing and social proofing, *Contemporary Social Science,* 16:2, 241-255. https://doi.org/10.1080/21582041.2019.1569714

Lewandowsky, S., & Yesilada, M. (2021) Inoculating against the spread of Islamophobic and radical-Islamist disinformation. C*ognitive Research: Principles and Implications,* 6, 1-15. https://doi.org/10.1186/s41235-021-00323-z

Petersen, M., Osmundsen, M., & Arceneaux, K. (2023) The "Need for Chaos" and Motivations to Share Hostile Political Rumors. *American Political Science Review,* 1-20. https://doi.org/10.1017/S0003055422001447

Reuters (2022) The changing news habits and attitudes of younger audiences. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/young-audiences-news-media

Vosoughi, S., Roy, H. & Aral, S. (2018) The spread of true and false news online. *Science,* 359(6380), 1146-1151. https://www.science.org/doi/10.1126/science.aap9559

Weikmann, T. & Lecheler, S. (2022) Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society.* https://doi.org/10.1177/14614448221141648

Wu, V., Carey, J., Nyhan, B., & Reifler, J. (2022) Legislator criticism of a candidate's conspiracy beliefs reduces support for the conspiracy but not the candidate: Evidence from Marjorie Taylor Greene and QAnon. *Harvard Kennedy School (HKS) Misinformation Review,* 3(5). https://doi.org/10.37016/mr-2020-103

Yesilada, M. & Lewandowsky, S. (2022) Systematic review: YouTube recommendations and problematic content. *Internet Policy Review.* https://doi.org/10.14763/2022.1.1652

CENTRE FOR RESEARCH AND EVIDENCE ON SECURITY THREATS

CREST Security Review provides a gateway to the very best knowledge and expertise. Its articles translate academic jargon to 'so what' answers and illustrate how behavioural and social science can be used effectively in everyday scenarios.

THE CENTRE FOR RESEARCH AND EVIDENCE ON SECURITY THREATS

CSR is produced by the Centre for Research and Evidence on Security Threats (CREST). CREST is funded by the UK's Home Office and security and intelligence agencies to identify and produce social science that enhances their understanding of security threats and capacity to counter them. CREST also receives funding from its core partners (the universities of Bath, Lancaster and Portsmouth). Its funding is administered by the Economic and Social Research Council (ESRC Award ES/V002775/1), one of seven UK Research Councils, which direct taxpayers' money towards academic research and training. The ESRC ensures the academic independence and rigour of CREST's work.

CREST has established a growing international network of over 140 researchers, commissioned research in priority areas, and begun to tackle some of the field's most pressing questions.

> '*CREST Security Review is a fantastic means by which we can keep practitioners, policy-makers and other stakeholders up-to-date on the impressive social and behavioural science occurring not only at CREST, but around the world.*'
>
> Professor Stacey Conchie, CREST Director

For more information on CREST and its work visit
www.crestresearch.ac.uk or find us on Twitter, @crest_research