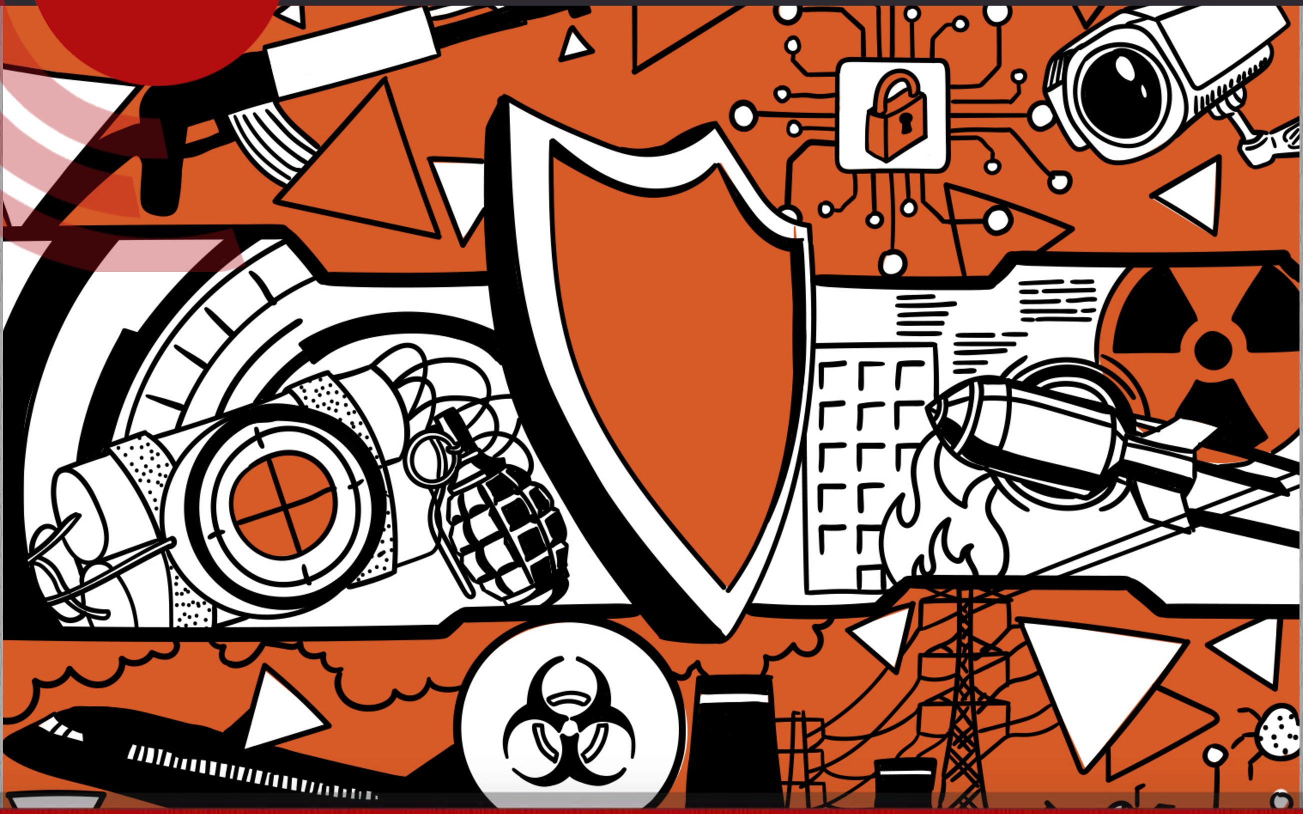
CREST Centre for Research and Evidence on Security Threats



Online Radicalisation:

A Rapid Review of the Literature

FULL REPORT

JULY 2023

ROSAMUND MUTTON
JAMES LEWIS
SARAH MARSDEN

JULY 2023

Online Radicalisation:
A Rapid Review of the Literature

FULL REPORT

Rosamund Mutton | University of St Andrews
James Lewis | University of St Andrews
Sarah Marsden | University of St Andrews

This report was produced as part of a CREST project to inform the refresh of the CONTEST strategy. The project provides updates to the evidence base behind key CONTEST topics. You can find all the outputs from this project at: www.crestresearch.ac.uk/projects/contest

ABOUT CREST

The Centre for Research and Evidence on Security Threats (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. Its funding is administered by the Economic and Social Research Council (ESRC Award ES/V002775/1).

www.crestresearch.ac.uk

Economic and Social Research Council

TABLE OF CONTENTS

EXECUTIVE SUMMARY

OVERVIEW

This guide sets out the evidence base for 'online radicalisation', examining how individual use of the Internet, in conjunction with offline influences, can facilitate radicalisation processes. The UK is the main context of concern, however comparable evidence is found in studies with samples from the USA, Canada, Belgium, Germany, Austria, and Israel.

Radicalisation remains a contentious concept and few studies explicitly define 'online radicalisation'. For the purposes of this guide, 'radicalisation' is understood as leading to cognitive outcomes reflected in changes in beliefs and ideas, and/or behavioural outcomes which manifest in changes in behaviour.

METHODOLOGY

Two systematic literature reviews (Hassan et al., 2018; Carthy et al., 2020) directed initial searches for relevant research. Further literature was identified through forward and backward citation searching, and narrower key word searches conducted in Google Scholar. Literature searches were completed between June and August 2022. The guide primarily examines literature published between January 2017 and July 2022. Although the evidence base remains modest in size, the research underpinning this guide is assessed to be good quality. There is a growing body of evidence that uses qualitative and quantitative methods to examine a range of factors which are relevant to online radicalisation.

KEY FINDINGS

- Online and offline activities and domains interact, challenging the 'online/offline dichotomy' popular in early research into online radicalisation. Radicalisation processes rarely take place in either the online domain or the offline sphere exclusively, but instead are characterised by complex and dynamic interactions between the two.

- Research that sought to distinguish between online and offline processes may have over-estimated the extent to which the Internet contributes to radicalisation processes. This tendency to focus on the role of the Internet may have come at the expense of recognising the role of offline factors and the importance of the interaction between online and offline contexts.

- The Internet in isolation does not cause radicalisation and is better understood as playing a role in facilitating this process. While the Internet can contribute to an individual's radicalisation, it cannot drive the process on its own.

BEHAVIOURAL RADICALISATION

- Use of the Internet can enable behavioural outcomes including event planning and preparatory activities, communication and networking behaviours (including arranging offline activities) and ideology-seeking actions.

- Pathways into violent extremism have been characterised as primarily offline, mainly online, and hybrid. Hybrid pathways seem to be the most common.

- There is no single profile of, or standard trajectory taken by, individuals whose use of the Internet influenced their radicalisation. However different pathways seem to be associated with differing levels of intent, capability, and engagement. Hybrid pathways demonstrate greatest engagement and intent; offline pathways, greatest capability; and online, the lowest levels of engagement, intent and capability.

COGNITIVE RADICALISATION

- Empirical research analysing the influence of online interactions and exposure to extremist content on violent extremist behaviour remains limited.
- Video-sharing platforms and social networking sites are spaces where individuals are most likely to encounter extremist content online.
- The individual is an active rather than passive actor in the radicalisation process. It is the individual's behaviour and how they utilise the Internet that informs its relevance to radicalisation.
- There is little robust evidence about whether and how recruiters try to identify or engage with those seeking out online extremist material.
- Individuals who actively seek out violent extremist material online seem to be at greater risk of radicalising and engaging in violence, compared to passive consumers.
- Research on the role exposure to violent extremist content online plays in cognitive radicalisation has suggested that initial exposure to extremist content online has the potential to trigger an interest in extreme ideologies, and that exposure to content from a combination of online and offline spheres may be more influential than exposure via one or the other.
- The amount of time spent online and willingness to express political views on the Internet seem to be associated with greater exposure to extremist material.
- A study that looked at personality traits, specifically the role of empathy, hostility, and aggression, found that aggression may be more influential than exposure to extremist propaganda in influencing extremist cognitions. However, research on the dynamics of these processes remains limited.

ONLINE INDICATORS OF BEHAVIOURAL RADICALISATION

- Robust empirical evidence on how online activities might be used to identify individuals at risk of behavioural radicalisation is comparatively weak.
- There is some evidence that exposure to extremist content online has a stronger link to radicalisation in comparison with other kinds of media-related risk factors, such as different platforms, mediums (e.g., Internet, newspaper etc.), content, activities, and attitudes.
- Recruiters may use different kinds of online extremist material to first nurture cognitive radicalisation and then try and move people towards violence.
- Some research suggests that posting patterns on social media may be able to differentiate between violent and non-violent extremists, and between behavioural and cognitive outcomes, but further research is needed to fully understand these processes.
- Future research is likely to benefit from combining computational and social science methods, and developing robust, publicly available standardised datasets which are free from bias.

INTERVENTION STRATEGIES

- The effectiveness of counter-narratives varies according to the intervention technique used and the type of outcome targeted.
- There is insufficient evidence to determine whether counter-narratives can prevent violence, however they may be able to address some of the risk factors associated with radicalisation.

- Inoculation theory may provide a foundation for developing deterrence strategies. This approach introduces individuals to weakened versions of an argument whilst providing evidence to refute it. Preliminary experiments indicate that 'active' inoculation methods (where the individual actively engages in a task such as a computer game) can improve critical thinking skills and reduce vulnerability to radicalisation. This research is at an early stage that will benefit from more attention before the potential risks, implications and scalability of this approach is understood.
- Although the evidence base is very limited, interventions may benefit from adopting a fine-grained approach that is tailored to specific audiences and online contexts, including audience segmentation and micro-targeting.
- Interventions have the potential to produce unintended outcomes, including further entrenching extremist views, for example where activists initiate arguments in response to extremist positions.
- There is some, limited evidence to suggest that highlighting the personal impact of involvement in extremism may be more effective than challenging extremist ideas or arguments, and that online interventions may be less effective with those with more entrenched views.
- Intervention providers working online will benefit from training and support to mitigate the risks associated with this work, and to ensure their approach is evidence-informed.

CHALLENGES TO UNDERSTANDING ONLINE RADICALISATION

- Accessing and gathering valid empirical data is one of the main barriers to producing robust research able to evidence whether, and to what extent, online activity influences violent offline behaviour. Similar difficulties arise in efforts to assess which factors influence attitudinal change.
- It can be difficult to generalise the findings of research drawn from small-n sample sizes collected using qualitative methods, or which focuses on a specific ideology or geographical context. Drawing broader conclusions to groups or settings beyond the data sample should be undertaken with caution.
- Large-n computational methods have the potential to identify broader trends in the data but can risk over-simplifying radicalisation processes.
- Efforts to understand the impact of online interventions face similar challenges to evaluations of offline P/CVE programmes. These include the difficulty understanding an intervention's impact; accessing appropriate data; ethical and security risks; and the difficulty identifying and evidencing the causal factors that shape outcomes.
- Methodological differences in how data are collected, used and analysed can be difficult to translate across disciplines.
- Ambiguous and/ or contested definitions of 'online radicalisation' can make it challenging to draw comparisons across studies which may be focused on different phenomena.

RECOMMENDATIONS FOR POLICY AND PRACTICE

- P/CVE interventions are likely to benefit from taking account of the hybrid nature of radicalisation processes and developing ways of targeting online and offline domains simultaneously, rather than separately. For example, by working in offline contexts to help develop digital literacy skills if the online space seems to be an important source of information for those engaged in primary or secondary interventions.
- Intervention strategies which provide an alternative source of meaning and association to replace the

relational networks offered by extremist groups, both online and offline, appear promising.

- There is some evidence to suggest it may be beneficial to prioritise interventions which focus on those who actively seek extremist content online, as they may be at greater risk of radicalisation to violence.
- The gamification (or use of mechanisms used in games) of interventions has the potential to appeal to those who actively seek extremist content. These types of intervention can encourage the development of critical thinking skills and may provide an element of interaction that active seekers are looking for.
- Interventions targeting video-sharing platforms and social networking sites may have a greater impact than targeting other areas online. However, there are risks to this approach. Counter-messaging videos and extremist content can share key words. This means that the algorithms which drive automated recommendation systems may direct users to extremist content, rather than to counter-messaging videos.
- Counter-narratives will benefit from careful targeting, taking account of the specific audience; the extent to which they may already be persuaded by extremist ideas; the risk factors the intervention is seeking to influence and the mechanisms by which positive outcomes might be enabled.
- Evidence regarding the impact of removing extremist content is limited. Taking down material may help to reduce its accessibility. However, there is some limited evidence that where material is removed from non-encrypted, more accessible online spaces, this has the potential to encourage users to move to encrypted platforms which are more difficult to monitor and moderate.
- Interventions should take account of unintended outcomes, including the potential to further entrench extremist views; generate risks to freedom of speech; and create incentives for tech companies to 'over-censor' content to avoid sanction.
- Intervention providers working online should receive appropriate training, professional development opportunities, and support.

DIRECTIONS FOR FUTURE RESEARCH

KEY AREAS OF FUTURE RESEARCH INCLUDE:

- Further work to understand the role of the Internet in pathways into extremism, including research able to interpret how online and offline dynamics interact.
- Research that draws on first-hand accounts of how the Internet shaped an individual's thinking and behaviour has the potential to elucidate the experiential aspects of radicalisation processes.
- Studies examining the impact of the COVID-19 pandemic on online radicalisation could try to assess the impact of lockdowns and whether associated feelings of isolation and the increased use of technology as a substitute for physical, face-to-face interactions led to greater exposure to, or engagement with, extremist content.
- Research which bridges computational approaches which analyse large amounts of data with social science-based methods able to interpret the experiential and subjective experiences of online users may provide greater insights and overcome the disjuncture between disciplines.
- Studies focused on a specific ideology could be carried out with data on a different ideology. This would help to determine whether findings can be generalised or are ideologically specific, and whether targeted interventions would benefit from being tailored to specific ideologies.
- Further research into the role of individual personality traits, pre-existing beliefs and other

psychological factors that may shape responses to extremist content and radicalisation. This would help tailor and target interventions in ways which are appropriate for particular groups or individuals, and help to avoid unintended or negative outcomes.

- Areas where results are limited, mixed or inconclusive would benefit from further research. These include:

  a. The relationship between exposure to extremist content online and cognitive radicalisation.

  b. Approaches able to interpret whether patterns of online engagement have the potential to identify individuals at risk of cognitive or behavioural radicalisation.

- Further work to understand the impact of interventions is important, assessing:

  a. What effect the removal of online extremist content has, and what risks this strategy carries.

  b. The potential of realist evaluation to develop a better understanding of which counter-narrative interventions work, for whom, under what circumstances, and why.

  c. The unintended consequences of different kinds of intervention strategy, including direct engagement online; efforts to direct people to counter messages; and counter-narrative material.

1. INTRODUCTION

The role of the Internet in radicalisation processes has attracted increasing attention from researchers. However, 'online radicalisation' is a conceptually ambiguous term. A targeted review of the literature (n=43) found that only 21 per cent (n=9) of studies on this topic defined the meaning of 'online radicalisation' (Macdonald & Whittaker, 2019). Studies that did define the term interpreted it in a variety of ways.

Radicalisation also remains a contentious concept. For the purposes of this guide, 'radicalisation'[1] is understood as leading to cognitive[2] and/or behavioural outcomes. Cognitive radicalisation produces outcomes relating to changes in beliefs and ideas, while behavioural radicalisation leads to changes in behaviour, including but not limited to, perpetrating violence (Winter et al., 2020; Herath & Whittaker, 2021).

'Online radicalisation' is broadly understood as a process where the Internet facilitates the search for, access to, and engagement with extremist content and networks. Through this, and in combination with experiences offline, individuals may come to gradually adopt beliefs that justify violence which, in some cases, may be translated into involvement in violence. Online platforms relevant to this process include social media sites; websites and forums dedicated to specific groups or networks; video sharing and streaming sites.

Research into the topic of online radicalisation exists across many disciplines, including terrorism studies; information and communications technology studies; sociology, psychology, and linguistics. This interdisciplinarity helps to provide different perspectives, although differing conceptualisations of online radicalisation and contrasting theoretical starting points can make it challenging to compare the results of studies.

There is an abundance of research analysing the 'supply-side' of online radicalisation, typically focused on examining the content of extremist material, while the 'demand-side' or the means by which individuals engage with the Internet and the impact it may have on cognitive or behavioural radicalisation, remains understudied (Hawdon et al., 2019; Bastug et al., 2020).

Empirical research exploring the influence of online interactions and exposure to extremist content on violent radical behaviour on- or offline (i.e., behavioural outcomes) is also limited (Hassan et al., 2018; Meleagrou-Hitchens et al., 2017; Shortland et al., 2022). This lack of evidence may have contributed to a tendency to overstate the Internet's role in radicalisation processes (Kenyon et al., 2022).

Concerns surrounding online radicalisation have intensified following the COVID-19 pandemic. Although restrictions on movement instigated by national lockdowns may have contributed to a decline in the number of terrorist attacks in the West between 2020 and 2021, attempts have been made by extremists to exploit the effects created by government measures, such as 'isolation, increased online activity, and resentment over…lockdown' (Institute for Economics & Peace, 2022, p. 14). It is too soon to draw clear conclusions, but the impact of the pandemic should be borne in mind in future research (Kenyon et al., 2022).

[1] As this guide draws on interdisciplinary research, terms and definitions used by authors vary. For consistency, terms such as 'radicalisation', 'terrorism', 'extremism' are used in accordance with the original study, however, a differentiation has been made between cognitive and behavioural radicalisation to aid the guide's structure.

[2] Sometimes referred to as 'attitudinal' outcomes.

2. OVERVIEW OF THE GUIDE

This guide sets out empirical evidence relating to online radicalisation in the UK, US, Canada, Belgium, Germany, Austria, and Israel published between January 2017 and July 2022. Three key studies published prior to 2017 were included as they were cited by many of the studies published post-2017 and are seminal in the field (von Behr et al., 2013; Koehler, 2014; and Pauwels & Schils, 2016). To be included in the guide, research had to be methodologically rigorous and informed by empirical evidence.

The identification of relevant literature was initially guided by two systematic literature reviews (Hassan et al., 2018; Carthy et al., 2020). Forward and backward citation searching, combined with narrower key word searches in Google Scholar were used to identify the studies included in the guide. Literature searches were completed between June and August 2022.

3. THE EVIDENCE BASE

This guide is based on open-source, published research only. It therefore does not benefit from research carried out by governments or civil society organisations that have not made their findings public.

A number of studies in this guide use mixed methods, typically analysing qualitative data using quantitative methods. In-depth and detailed primary data collected through interviews tends to be drawn from smaller samples.[3] Larger scale analyses of data draw on pre-existing, open access databases such as the Profiles of Individual Radicalization in the United States (PIRUS), or the Global Terrorism Database (GTD),[4] and a limited number of observational experiments.[5] The largest sample sizes tend to be drawn from survey or questionnaire responses.[6]

Although the size of the evidence base remains modest, the strength of the research included in this guide is assessed as good. It includes a number of methodologically robust quantitative studies, which set out in detail the type of data gathered, the methods of analysis, and the limitations of the research.

Two systematic reviews found the number of studies using empirical evidence was substantially smaller than the number of studies on the topic of online radicalisation (Hassan et al., 2018; Carthy et al., 2020) and there have been repeated calls for more empirical research in this area (e.g., Gill et al., 2017; Conway, 2017; Odağ et al., 2019; Scrivens et al., 2020). A key challenge is summarised by Herath & Whittaker (2021) who note the 'paucity of empirically valid research into how the Internet impacts pathways towards terrorism' (p. 3).

Empirical research pre-dating 2017 on online radicalisation has substantial gaps. Wolfowicz et al., (2022) and Gaikwad et al., (2021) found that a proportion of this research used datasets which contained sources of bias, while Hassan et al. (2018) noted the difficulty in finding high-quality, empirically robust studies. However, the evidence base is growing, and a number of more robust empirical studies have been published since Hassan et al.'s (2018) review (e.g., Kenyon et al., 2022; Herath & Whittaker, 2021; Saleh et al., 2021; Gaudette et al., 2020; Baugut & Neumann, 2020).

Although the research designs and methodological approaches in the research in this report are strong, the evidence base has a number of limitations. A common challenge acknowledged by many studies is the difficulty generalising their findings beyond the sample demographic. Most findings are specific to the data sample, ideological context, and socio-political and cultural setting where they were conducted. Where studies do draw tentative broader conclusions, they urge caution against making decontextualised generalisations.

[3] For example: von Behr et al. (2013) [n=15]; Koehler (2014) [n=8]; Gaudette et al. (2020) [n=10]; and Baugut & Neumann (2020) [n=44].

[4] For example: Herath & Whittaker (2021) [n=231] and Youngblood (2020) [n=416]. Both studies draw on biographical data, as does Mills et al. (2020), although this study utilises a much smaller dataset [n=4].

[5] See Saleh et al. (2020) [n=291]; Lewandowsky & Yesilada (2021) [n=591]; and Braddock (2022) [n=357].

[6] For example: Frissen (2021) [n=1,872]; Pauwels & Schils (2016) [n=6,020]; Shortland & McGarry (2022) [n=479]; and Hawdon et al. (2019) [n=768].

4. ANALYSIS

4.1. OVERVIEW

The analysis that follows sets out the research on online radicalisation, describing the current evidence about the Internet's role in these processes. Section 4.2. considers research on the behavioural outcomes of radicalisation, including studies which have outlined the different pathways that individuals can take towards violent extremism, and the role the Internet may have played in this process. Section 4.3. examines the evidence base surrounding cognitive radicalisation. Two key themes in this literature are the distinction between 'active' and 'passive' seekers of extremist material, and the limited knowledge there is about the impact of exposure to extremist content on cognitive radicalisation.

Section 4.4 reviews research that has sought to identify online indicators of behavioural radicalisation and considers the implications of this limited evidence base for efforts to identify individuals who may be at risk of transitioning from non-violent to violent behaviour. Section 4.5. outlines research on intervention strategies before an overview of common barriers and challenges facing the research in Section 4.6. is given.

4.2. BEHAVIOURAL RADICALISATION

This section outlines the literature on behavioural radicalisation concerned with violent and non-violent behavioural outcomes. This research emphasises the importance of considering behaviour, not just beliefs, when trying to understand radicalisation processes (e.g., Gill et al., 2017). Importantly, behaviours do not necessarily have to be violent. Owing to the difficulty in accessing robust, empirical data which explicitly links online activity to violent behaviour offline, studies also analyse non-violent activities such as distributing propaganda or communicating with like-minded peers.

A number of studies draw on data from individuals convicted of terrorist or extremist offences to understand the processes associated with behavioural radicalisation (Gill et al., 2017; Herath & Whittaker, 2021; Gaudette et al., 2020; Kenyon et al., 2022). Because of the range of offences this covers, these studies capture a variety of violent and non-violent behaviours.

Research identifying different pathways towards radicalisation explores the Internet's role in shaping behavioural outcomes. Two studies examine the process and impacts of online, offline and hybrid dynamics in relation to different levels of engagement with extremist content; intent to commit violent extremism offences; and capability to carry out such offences (Kenyon et al., 2022; Herath & Whittaker, 2021).

In line with the UK Government's definition (HM Government, 2012), 'intent' is interpreted as readiness to commit extremist violence. 'Capability' refers to whether an individual is realistically able to commit this violence, for example, whether they have access to the required materials and expertise necessary to plan and to conduct an attack. Whilst 'engagement' reflects the motivations, needs and influences that shape pathways into extremism.

Key Findings

- Online and offline activities and domains interact, challenging the 'online/offline dichotomy' popular in early research into online radicalisation. Radicalisation processes rarely take place in either the online domain or the offline sphere exclusively, but are characterised by complex and dynamic interactions between the two.
- Research that sought to distinguish between online and offline processes may have over-estimated the extent to which the Internet contributes to radicalisation processes. This tendency to focus on the role of the Internet may have come at the expense of recognising the role of offline factors.
- The Internet in isolation does not cause radicalisation and is better understood as playing a role in facilitating this process. While the Internet can contribute to an individual's radicalisation, it cannot drive the process on its own.
- Use of the Internet can enable behavioural outcomes including event planning and preparatory activities, communication, and networking behaviours (including arranging offline activities) and ideology-seeking actions.
- Pathways into violent extremism have been characterised as primarily offline, mainly online, and hybrid. Hybrid pathways seem to be the most common.
- There is no single profile of, or standard trajectory taken by, individuals whose use of the Internet influenced their radicalisation. However different pathways seem to be associated with differing levels of intent, capability, and engagement. Hybrid pathways demonstrate greatest engagement and intent; offline pathways, greatest capability; and online, the lowest levels of engagement, intent and capability.

4.2.1. ONLINE-OFFLINE INTERACTIONS IN BEHAVIOURAL RADICALISATION

Online and offline domains interact. Individuals engage in a range of activities which span both environments (Gill et al., 2017; Herath & Whittaker, 2021; Gaudette et al., 2020; Baugut & Neumann, 2020). This suggests that radicalisation is experienced as a hybrid process involving online and offline processes.

The Internet can help build capacity to support attack planning. From a database of 223 UK convicted terrorists constructed by Gill et al., (2017), evidence of online activity relating to actors' radicalisation and/or attack planning was found in 61 per cent of cases. The database disaggregated types of online behaviour. A third (32%) of the 61 per cent of cases prepared for their attack by accessing and using online resources, such as bomb-making instruction videos; body disposal techniques; and plans of transport networks, illustrating the potential role online activities can play in attack preparation.

This study also found that 29 per cent communicated with other radicals virtually, and at least 30 per cent accessed extremist ideological content online, although it was noted that in some cases, because of the scale of material accessed, it may have been unlikely that individuals consumed and understood it all in full (Gill et al., 2017).[7]

[7] One perpetrator had downloaded 17,779 computer files of ideological material, 1,152 of which contained extremist content (Gill et al., 2017, p 107-108).

The Internet has also been used as a tool to facilitate and organise events and social interactions offline, such as attending events held by extremist groups or meeting online contacts in person. Three of the 10 Canadian former right-wing extremists interviewed by Gaudette et al. (2020) said they had used the Internet for this purpose during their involvement in violent racist skinhead groups.

The authors concluded that 'the Internet can serve as a gateway for individuals to engage in violent extremist activities offline, connecting adherents in the online world to the offline world, oftentimes through the online promotion of offline events (e.g., concerts, rallies, protests, and gatherings).' (Gaudette et al., 2020, p. 13). Owing to the range of factors involved, and the small sample size, it is not possible to assess whether the radicalisation of individuals in Gaudette et al.'s study was dependent on the Internet. However, there is evidence that the immersive function performed by the online space, and the way it enables individual access to extremist networks and content, was an important mechanism which seemed to contribute to many of the participants' radicalisation processes (Gaudette et al., 2020, p. 13).

4.2.2. PATHWAYS INTO VIOLENT EXTREMISM

Two retrospective analyses of extremist offenders' life histories have identified different types of radicalisation pathway (Kenyon et al., 2022; Herath & Whittaker, 2021).[8] This research found that the Internet is used to a greater or lesser extent according to the type of radicalisation pathway taken by an individual, and provides further evidence for the intertwining of offline and online domains in many cases of radicalisation. Table 1 offers a simplified comparison of pathways towards radicalisation identified in these two studies.

Herath, & Whittaker (2021). Online Radicalisation

Four behavioural radicalisation pathways were identified through an analysis of antecedent behaviours - including online and offline networking and event preparation - prior to attempted terrorist attacks by US-based Islamic State actors charged with terrorism offences (n=231) (Herath & Whittaker, 2021). The pathways varied according to the level of engagement with online and/or offline domains and demonstrate the complex and dynamic interaction between the two. As well as analysing the role of variables specific to online and offline contexts, the study found correlations between equivalent online and offline

	Herath & Whittaker (2021)	Kenyon et al. (2022)
'Pathway' to radicalisation	Integrated: n=103 [hybrid online and offline factors]	Hybrid: n=113 [combination of online & offline influence]
	Encouraged: n=63 [primarily online, but still engaged in offline]	'Internet' group: n=29 [primarily online]
	Isolated: n=38 [lack of interaction both online and offline]	---------
	Enclosed: n=27 [primarily offline, but still engaged in online]	'Face-to-face' group: n=93 [primarily offline]

Table 1. Radicalisation Pathways Identified in the Literature

[8] A slightly earlier study helped to establish understanding radicalisation processes as 'pathways' by analysing life narratives (n=56) of violent (n=31) and non-violent (n=25) radicalised individuals in the US (Jensen et al., 2020). While this study did not explore the role of the Internet, it did provide empirical evidence that radicalisation to violence is a much more complex and multi-faceted set of processes than previously hypothesised (p 1083-1084). Its findings aided movement away from simplified conceptualisations of linear radicalisation processes.

behaviours, indicating that actors used both environments to carry out comparable activities.[9]

1. **Integrated**

This pathway represents the experiences of individuals who demonstrated a high level of engagement in both online and offline domains and who were part of a wider network when planning their attack. Those involved in offline networks were also in contact with individuals online beyond their in-person conspirators.

2. **Encouraged**

Individuals on this pathway made significant use of the Internet whilst networking and planning their attack and had less in-person contact with co-ideologues. Although their wider network was predominately online, the authors caution that 'the dynamics behind terrorist behavioural pathways are deeply complex and can no longer be conceptualised as simply online or offline, but some degree of both.' (p.12), even in pathways where the Internet played a more significant role.

3. **Isolated**

Actors categorised as 'isolated' had comparatively limited interaction with co-ideologues either online or offline. Interactions did take place, but these were at a much lower level than with the other three pathways. Individuals differed as to whether they used the Internet as an event planning tool.

4. **Enclosed**

The 'enclosed' pathway describes the experiences of those whose network activity was greatest in offline domains, but who made use of the Internet to plan their attack. While these individuals typically had stronger offline ties to a small peer group, this did not preclude other members of this group from engaging with a wider network of co-ideologues online. In these cases, they could mediate between their peer group and extremists online.

Kenyon, et al. (2022). Online Radicalization

Analysis of a sample of convicted extremist offenders (n=235) in England and Wales identified three radicalisation pathways (Kenyon et al., 2022). These pathways were based on the relevance of Internet use to the radicalisation process. Actors who primarily radicalised online were allocated to the 'Internet' group (n=29); those who mainly radicalised offline were categorised as the 'Face-to-face' group (n=93); and those who radicalised through a combination of online and offline influences were placed in the 'Hybrid' group (n=113).

This study found that different radicalisation pathways were associated with differing levels of engagement, intent, and capability to commit a terrorist offence (see Table 2) (Kenyon et al., 2022). The hybrid group had the highest levels of engagement with an extremist cause and intent to commit an offence. Those who primarily radicalised offline had the highest overall levels of capability, whilst those in the Internet group had the lowest levels of engagement, intent and capability.

The authors concluded that 'contact with other extremists in an offline setting plays a crucial role in moving individuals from holding extremist views and taking an interest in a specific group or cause, to a desire to act on behalf of that group or cause' (Kenyon et al., 2022 p. 12). This relationship between offline social connections and levels of intent was reflected in the majority, although not all, of the cases across the sample.

Other factors that Kenyon et al (2022) found were statistically significant include:

[9] Networking behaviour variables included: 'maintained contact with a network online'; 'recruited others offline'; 'attended a wider network event'; and 'sought legitimisation offline'. Event behaviour variables included: 'learned and planned offline'; 'overcame hurdles online'; 'experienced a motivating factor offline'; and 'online financial transaction'. Online and offline equivalents for some behaviours (e.g., 'disseminated propaganda online' and 'disseminated propaganda offline') were included.

- Age: Those in the hybrid group were more likely to be under 25 compared to those in the offline group.
- Prior offending: Individuals who had primarily radicalised offline were more likely to have a history of criminality compared to the hybrid group.
- Mental illness/ personality disorder: Those whose radicalisation was primarily online were more likely to have a mental illness/ personality disorder 'strongly present' in their risk assessments in comparison to both the hybrid and offline groups.
- Violent/ non-violent offence: The face-to-face group were more likely to have been convicted for a violent offence than those in the online group and the hybrid group, although the relationship was weaker with the latter group.
- Ideology: Those who primarily radicalised online, and those in the hybrid group were both more likely to be committed to Islamist extremism than the face-to-face group.

Primary arena of radicalisation	Engagement with an extremist group / cause	Intent to commit extremist offences	Capability to commit extremist offences
Online	Lowest	Lowest	Lowest
Offline	Mid	Mid	Highest
Hybrid	Highest	Highest	Mid

Table 2. Association Between Pathways and Engagement, Intent and Capability (Kenyon et al., 2022)

4.2.3. CONCLUSION

Research seeking to identify different pathways towards behavioural radicalisation demonstrates the differing role the Internet can play. There is growing evidence that radicalisation tends to occur in hybrid environments, where online and offline influences interact. This challenges the idea of purely online forms of radicalisation. While online influences can contribute to radicalisation processes, they do not operate in a vacuum and are often influenced by offline factors. Radicalisation processes have therefore been described as 'cyber-enabled' rather than 'cyber-dependent' (Gill et al., 2017, p. 114), and as a potential 'gateway' to further radicalisation (Gaudette et al., 2020, p. 13).

4.3. COGNITIVE RADICALISATION

The Internet has increased the amount of extremist content that is available in ways which are often viewed as a mechanism able to facilitate the development and advancement of extremist worldviews (Koehler, 2014). The most common spaces where individuals encounter extremist material are video-sharing and social networking sites (Baugut & Neumann, 2020; Nienierza et al., 2021). However, empirical evidence about the impact of consuming this material is limited. This section examines research which considers the type of content, (for example e-magazine; video; chat forum etc.); whether it is actively sought out; and how it interacts with other individual-level characteristics influence cognitive radicalisation.

Key Findings

- Empirical research analysing the influence of online interactions and exposure to extremist content on violent extremist behaviour remains limited.
- Video-sharing platforms and social networking sites are spaces where individuals are most likely to encounter extremist content online.
- The individual is an active rather than passive actor in the radicalisation process. It is the individual's behaviour and how they utilise the Internet that informs its relevance to radicalisation.
- There is little robust evidence about whether and how recruiters try to identify or engage with those seeking out online extremist material.
- Individuals who actively seek out violent extremist material online seem to be at greater risk of radicalising and engaging in violence, compared to passive consumers.
- Research on the role exposure to violent extremist content online plays in cognitive radicalisation has suggested that initial exposure to extremist content online has the potential to trigger an interest in extreme ideologies, and that exposure to content from a combination of online and offline spheres may be more influential than exposure via one or the other.
- The amount of time spent online and willingness to express political views on the Internet seem to be associated with greater exposure to extremist material.
- Personality traits, specifically aggression, may be more influential than exposure to extremist propaganda in influencing extremist cognitions.

4.3.1. EXPOSURE TO VIOLENT EXTREMIST CONTENT ONLINE

Empirical studies on the effects of online exposure to extremist content on cognitive radicalisation have produced a range of findings. In some studies, initial exposure to extremist content online is reported to trigger an interest in extreme ideologies. Gaudette et al.'s (2020) interviews with ten former Canadian right-wing extremists illustrated that 'regardless of how they were first exposed to the violent extremist content online, these participants oftentimes described this exposure as a critical point that sparked their initial interest in violent extremist ideologies.' (p. 7).

A larger study conducted on people aged 15 – 24 in the US (n=494) suggests there may be a correlation between engaging in particular kinds of activities and exposure to violent extremist content (Costello et al., 2020). Readiness to express political views online and the amount of time per day spent online were correlated with greater exposure to online extremism. However, the authors were unable to determine how the rate of exposure to extremist material related to negative outcomes. This reflects a wider theme in the literature regarding the difficulty of interpreting the role online activities play in influencing offline violence.

Other studies suggest that the combined effect of being exposed to extremist material in online and offline domains may be more influential than exposure via either one alone (Kenyon et al., 2022). This conclusion is informed by research that found those who primarily radicalised via either the online or the offline sphere had lower overall levels of intent to commit extremist offences (Kenyon et al., 2022).

Research is beginning to examine the role individual characteristics play in the context of exposure to extremist content online. As part of a wider psychological study on human behaviour, one study surveyed a sample of the general population in the US aged between 18 to 26 to assess the impact of extremist content on extremist cognitions (Shortland et al., 2022). The justification for targeting participants within this age group was the authors' view that young adults are potentially more vulnerable to radicalisation. The study considered the role of trait-level empathy, aggression, and hostility (n=1,112) and concluded there was 'no consistent effect of the propaganda on extremist mindset' (Shortland et al., 2022, p.15). However, personality factors were found to have a more significant effect. Higher scores on extremist cognitions were linked to higher levels of trait aggression, while lower levels of extremist cognitions were linked to greater levels of empathy.

The benefits of understanding individual differences such as personality traits and existing attitudes were further supported by a study into the neurocognitive processes linked to online radicalisation (n=10) (Howard et al., 2022). This study examined the role of existing belief systems in shaping responses to extremist material. Where the content of messages was in line with pre-existing beliefs, the messages were able to generate empathy, which increased the persuasive impact of the material and generated 'radical-persuasive outcomes'. The study concluded that the 'pathway from message exposure to radicalization is heavily influenced by an individual's existing belief system and the capacity to reconcile one's belief system with the propaganda message through the process of empathy' (Howard et al., 2022, p. 4). Empathy with message content which reflects pre-existing beliefs could 'remove psychological resistance to violence' (ibid, p. 19) and in this way has the potential to contribute to radicalisation processes. This may help explain why extremist content does not radicalise all individuals who engage with it.

4.3.2. ACTIVELY SEEKING EXTREMIST MATERIAL ONLINE

Individuals who deliberately search for extremist material online (referred to as 'active seekers') appear to be at greater risk of engaging in violence than passive seekers (those who are accidentally exposed to or encounter such material) in offline settings (Hassan et al., 2018, p. 71; Frissen, 2021; Pauwels & Schils, 2016). One interpretation of this finding is that accessing online content may influence violent behaviour offline. Alternatively, those more persuaded by the benefits of violence may seek out this material. Further research is needed to unpick the causal relationships at work between actively searching for extremist content and sympathy and engagement in violence.

Only one study identified for this guide examined the relationship between actively seeking different forms of online extremist propaganda and radicalisation outcomes and is set out in the box on the next page.

4.3.3. CONCLUSION

There is currently insufficient empirical evidence to draw firm conclusions about what shapes cognitive radicalisation. Research to date has focused on what shapes exposure to extremist material and whether it was actively or passively sought out. Preliminary research analysing jihadist information seeking suggests that those who actively seek (e-)magazine content are likely to score higher on radicalisation measures. This contrasts with those who search for audio-visual material which, while more sought-after, was least predictive of sympathy for violent radicalisation.

There is a small body of research that examines the role of individual personality traits and pre-existing beliefs in cognitive radicalisation. This suggests these factors may be influential in shaping pathways into extremism but more research is needed to understand the relationships between individual characteristics, online extremist content, and radicalisation processes.

Case Study

THE LINK BETWEEN ACTIVELY SEEKING JIHADIST INFORMATION AND COGNITIVE RADICALISATION IN BELGIUM

Frissen (2021) examined what influenced cognitive radicalisation by looking at the nature of the information seeking process; the role of moral disengagement; prior involvement in petty crime; and socio-demographic data.

Based on a sample of 1,872 Belgian young adult responses to a questionnaire, Frissen found that the strongest direct predictor for cognitive radicalisation was moral disengagement. The author argued that moral reasoning plays a key role in radicalisation processes, suggesting that 'jihadist information seeking is associated with cognitive radicalisation, through a cognitive process of moral disengagement' (p. 9). The interaction between jihadist information seeking, moral disengagement and juvenile delinquency was found to predict 'almost 50% of an individual's cognitive radicalization [sic]' (p. 9).

Audio-visual material of beheadings was the most sought-after material by respondents: a little over a third (36%) actively sought out this content. However, actively searching for audio-visual material was found to be the 'least predictive for sympathies for violent radical behaviours' (p. 8). Whereas the 10-11 per cent who sought out static (e-)magazine content scored 'significantly and substantially higher on the radicalisation scale' than those who did not seek this type of material (Frissen, 2021, p. 8).

Frissen speculates that the difference in popularity between audio-visual materials and static magazines, and the association of static magazines with higher levels of radicalisation, relates to the intended purposes of these types of material. Citing his previous research as evidence (including Frissen & d'Haenens (2017); Frissen, Toguslu, Van Ostaeyen, & d'Haenens (2018)), Frissen argues that jihadist magazines use 'psychological and dogmatic rhetoric with the sole purpose to recruit, inspire, and radicalize [sic] their audiences' (p. 9). In contrast, audio-visual material is designed to attract public attention. From this, Frissen tentatively suggests that each type of material does what they are 'designed for'. This study also highlights the insights that are possible when the role, appeal, and impact of alternative kinds of online content are differentiated.

4.4. ONLINE INDICATORS OF BEHAVIOURAL RADICALISATION

This section reviews research that has sought to identify links between online and offline behaviours. Overall, there is a lack of robust empirical research assessing whether it is possible to identify individuals moving towards behavioural radicalisation based on patterns of online behaviour. However, the field has made some preliminary efforts to explore whether online engagement, including a user's language, can be used to interpret the move from non-violent to violent behaviour. This has implications for the design of CVE interventions and the development of measures for detecting those at risk of behavioural radicalisation (reviewed in section 4.5.).

Key Findings

- Robust empirical evidence on how online activities might be used to identify individuals at risk of behavioural radicalisation is comparatively weak.
- There is some evidence that exposure to online extremist content has a stronger link to radicalisation in comparison with other kinds of media-related risk factors such as different platforms, mediums (e.g., Internet, newspaper etc.), content, activities, and attitudes.
- Recruiters may use different kinds of online extremist material to first nurture cognitive radicalisation and then try and move people towards violence.
- Some research suggests that posting patterns on social media may be able to differentiate between violent and non-violent extremists, and between behavioural and cognitive outcomes, but further research is needed to fully understand these processes.
- Future research is likely to benefit from combining computational and social science methods, and developing robust, publicly available standardised datasets which are free from bias.

A systematic literature review of 53 studies analysed the effects of 23 media-related risk factors, including different mediums (e.g., Internet, newspaper, radio etc.); platforms (e.g., Facebook, Twitter); content (e.g., violent, general etc.); activities (e.g., posting, consuming etc.); attitudes (e.g., network attachment, perception of bias); and other individual factors (e.g., technical skills etc.), on cognitive and behavioural radicalisation (Wolfowicz et al., 2022). Although the authors caution that the quality of the evidence they assessed was low and more robust study designs are needed, they reached the following conclusions:

- 'Simple media consumption is unlikely to be associated with any significant risk of radicalization [sic]'

- 'Internet-mediated exposure to radical content, whether passive or active, is associated with a significantly stronger relationship with radicalization than other types of media-related risk factors.' (Wolfowicz et al., 2022, p.3)

By focusing on engagement with different types of extremist content, a study by Baugut and Neumann (2020) (n=44) examined the behavioural and cognitive radicalisation processes of those who became either violent or non-violent radical Islamists. Interviews revealed that as an individual moved from cognitive to behavioural radicalisation, they engaged with different types of extremist content, influenced by those seeking to recruit them.

During cognitive radicalisation content focused on religious themes, whereas during behavioural radicalisation, material shifted to the role of violence. Beginning with content focused on victimhood, respondents reported consuming material calling for revenge and exhorting them to support the caliphate, before engaging with 'apocalyptic propaganda' underlining the importance and urgency of the need for violence. Finally, they turned to instructional material that provided them with 'information, distraction, and escapism' (Baugut & Neumann, 2020, p.1586).

4.4.1. COGNITIVE OUTCOMES

A growing body of research examines the cognitive outcomes of engagement with online extremist content using computational scientific approaches (e.g., Araque & Iglesias, 2022; Lara-Cabrera et al., 2019). Research falls into three broad categories: analysis of online radicalisation processes; detection of radical content and users online; and prediction of radicalisation (Fernandez et al., 2021).

These studies are characterised by their use of artificial intelligence (AI) technologies to analyse large amounts of data. A common study design involves using machine learning techniques to conduct searches for key terms considered to be indicative of radicalisation (e.g., Badawy & Ferrara, 2018). Studies focus on classifying and analysing the semantic and lexical content used by individuals declaring support for extremist groups online.

This research varies in its methodological and empirical rigour. One systematic review of the literature (n=64) found that online detection of cognitive radicalisation is limited by several factors (Gaikwad et al., 2021):

- Datasets were found to contain bias because they were produced by studies disproportionately analysing data on Islamist ideologies. Developing classification algorithms to identify indicators of online radicalisation that are predominately trialled on Islamist extremist data may mean they are less able to detect online radicalisation using data on other ideologies.

- Many of the studies analysed data which was described as 'class-imbalanced', meaning that the number of comparative samples within one classification category outnumbers those allocated to other classification categories (Tharwat, 2021). The machine learning algorithms used in the reviewed studies, assume that the data is balanced (i.e., comprises a similar number of samples within each classification category) and produces analyses based on this assumption. This method may be less appropriate for data which is subjective, such as that used to detect online radicalisation. Multiple interpretations of this data can be made, resulting in the number of samples within one classification category outnumbering those classified as other categories and producing class imbalance.

The empirical evidence is weakened by the quality of the data used. The construction of custom datasets is a common method to overcome data collection challenges. This involves collating data from multiple sources to build a dataset specific to a certain project. Where these use subjective validation methods over statistical techniques, custom datasets may include data that has not been robustly verified or validated. Publicly accessible, standardised datasets would help to ensure the quality of the data analysed (Gaikwad et al., 2021; Fernandez et al., 2019).

Criticisms have been made of research using data mining[10] and natural language processing techniques[11] because of the way some of them classify the data. Research seeking to detect online extremist material typically groups data into discrete categories e.g., 'extremist–non-extremist' or 'radical–non-radical'. Classifying individuals as either 'extremists' or 'non-extremists' according to their online behaviour has been described as over-simplifying the radicalisation process and failing to recognise the individualised nature of these experiences (Ajala et al., 2022; El Barachi et al., 2022). These methods are also unable to capture offline influences such as the role of peers and social networks. Although able to provide some insights, these approaches are therefore less able to take account of the shift towards understanding radicalisation as a hybrid process, spanning both online and offline spheres.

Research which combines theories and methods from the social sciences, such as lexicon or discourse analyses, to analyse data gathered using computational approaches, for example through data scraping, have the potential to inform broader conclusions (e.g., Fernandez et al., 2019; Lara-Cabrera et al., 2017; El Barachi et al., 2022).

Research by Fernandez et al. (2019) uses an interdisciplinary approach to understand online influence. The study applies the 'roots of radicalisation' theory to a dataset gathered from Twitter (n=224) to create and test the effectiveness of an algorithm to detect and predict radicalisation influence, before analysing 112 pro-Islamic State (IS) accounts to understand their social influence. Beginning with a theory of radicalisation meant it was understood as an experiential process, capturing interactions between micro, meso, and macro levels, rather than adopting a binary classification approach to the data.

The study found that while 'general' and pro-IS users often reported on the same current events, using the same key terms, there were important differences in tone and message portrayal. Pro-IS users framed their posts with a propagandistic tone (Fernandez et al., 2019). Binary classifications would likely categorise content by both user groups as evidence of signs of radicalisation. However, by applying a social science theory it made it possible to distinguish between different levels of support for IS. This distinction is key to targeting interventions to specific audiences.

Another study used five factors relating to personality, attitudes, and beliefs that were identified by experts as radicalisation indicators to assess the risk of radicalisation in social networks on Twitter (Lara-Cabrera et al., 2019). Users who tended to write longer posts which expressed negative sentiments, and used more swear words, were more likely to be radicalised or considered at greater risk of radicalisation. Other radicalisation indicators included discussing Jihadism in a positive manner, being negative about Western society, and describing perceptions of being discriminated against (Lara-Cabrera et al., 2019).

One of the few studies to analyse far-right extremism drew on a dataset of 259,000 tweets and found that current events had a significant impact on users expressing negative sentiments (El Barachi et al., 2022). User behaviour over time was analysed through a range of social and computer science approaches, including sentiment, emotion and social circle analysis. While the sharing of negative sentiments fluctuated, increases tended to be a 'temporary reaction to political events or attacks on political opponents' (El Barachi et al., 2022, p. 203).

[10] Interesting trends and patterns within data are identified through data mining processes utilising algorithms (Roiger, 2017).
[11] Humans understand language by processing it. Natural language processing techniques utilise human understanding and use of language to develop computer systems' understanding and manipulation of natural language text or speech to perform various tasks (Nagy, 2018).

These studies illustrate the benefits of combining theories and methods from different disciplines to provide more contextualised and nuanced understandings of large datasets.

4.4.2. BEHAVIOURAL OUTCOMES

In response to the need for robust empirical studies able to interpret how online behaviour relates to violent behaviour offline, a small number of experimental studies have begun to derive findings that might help identify those at risk of further radicalisation. These studies are methodologically sophisticated; however this remains a new area of research and the conclusions will benefit from further investigation and verification.

Experimental testing of radicalisation frameworks has produced inconclusive findings, and has not been able to reliably detect whether an individual is in the process of radicalising towards violence. Research which developed and tested the Ontological Framework to Facilitate Early Detection of 'Radicalization' [OFEDR] risk model found it was only able to offer a low probability of the early detection of radicalisation. However, this was still higher than detecting the probability of a terrorist act. The author concluded that the gap between the two probabilities could represent a 'region of interest in the work of preventing escalation of radicalisation processes' (Wendelberg, 2021, p. 22), suggesting an area for future research.

More promising findings emerged from a study which applied a two-stage framework to analyse just under 37,000 online posts (Theodosiadou et al., 2021). It sought to identify points in time where a change in attitude towards terrorism and/or participating in activities related to terrorism could suggest escalation in radicalisation. Following the classification of online text as terrorism or hate-speech related, change point detection (CPD) algorithms were used to analyse the time series generated by the textual data.

Research examining posting patterns on social media has found differences between violent and non-violent extremists. One study of right-wing extremists (RWEs) on Stormfront Canada (n=99) found differences in the frequency and overall amount of online activity between violent and non-violent individuals (Scrivens et al., 2021). Violent RWEs tended to be much less active online than non-violent RWEs (Scrivens et al., 2021). The authors speculate that this may be because violent RWEs tend to be more clandestine. Their participation in violence may generate greater fear of detection by the authorities, and paranoia regarding keeping personal information and identities secret which leads them to limit their online behaviours.

An analysis of Facebook profiles of 48 lone actor terrorists who carried out ideologically or political motivated attacks in Israel between 2014-2018, found that 'specific sets of behaviours or patterns of activities may be more easily classified as indicators of the move from radical beliefs to radical behaviors [sic]' (Wolfowicz et al., 2021, p. 8). Examples of the metrics studied included the likelihood of posting about a terrorist attack perpetrated by a Facebook connection; the number of radical compared to non-radical posts; and whether posts were text- or image-based.

The study found that shared posts made up almost a third of the terrorists' posts, and were more likely to be comprised of images. Non-violent radicals were more likely to author their own, text-based posts, although uploading images was still common. Specifically, 'the ratio between originally authored text-based posts and shared posts was a statistically significant predictor … [distinguishing] the non-violent radical and terrorist cases.' (Wolfowicz et al., 2021, p. 7). This research illustrates the potential for analyses of social media to distinguish between different kinds of users online, leading the authors to argue these factors could be applied to the offline domain, to predict offline behaviour. However, because this research is rooted in a specific geographic context, efforts to generalise the findings beyond this setting should be undertaken cautiously.

4.4.3. CONCLUSION

There is relatively little robust research assessing whether online engagement can detect a growing commitment to extremism and identify individuals at risk of radicalisation. The research that has been carried out provides tentative support for the argument that it is possible to identify patterns of online behaviour that may help interpret the potential for offline violence. However, the need for more research able to inform 'early warning systems' has been identified by various reviews of the literature. Given the potentially serious implications of efforts to detect and disrupt online activities, including the possibility of negative unintended consequences, this effort should proceed cautiously.

4.5. INTERVENTION STRATEGIES

This section describes research that has sought to understand the scope and effectiveness of different kinds of interventions to address online radicalisation. It begins by outlining research analysing the scope and effectiveness of counter-narratives which typically target cognitive radicalisation by providing information that attempts to undermine extremist propaganda and beliefs. A second intervention approach involves the removal of material from the Internet. There is only limited data by which to assess of the effectiveness of this strategy.

4.5.1. COUNTER-NARRATIVES

Although definitions vary, counter-narrative interventions are generally understood as 'narratives comprised of content that challenges the themes intrinsic to other narratives. In the context of CVE [countering violent extremism], counternarratives challenge themes within terrorist narratives that are consistent with the group's ideology' (Braddock & Horgan, 2016, p.386).

The most robust evidence on the impact of counter-narratives comes from a systematic review which identified 19 studies published between 2000 and 2018 (Carthy et al., 2020). From the range of counter-narrative approaches identified, just under half (48%) used counter-stereotypical exemplars that sought to challenge stereotypes and provide prosocial or moral exemplars to challenge the dominant narratives found in extremist discourse. Other approaches included alternative narratives and inoculation techniques. Their effectiveness varied according to the intervention technique and the type of outcome targeted.

Overall, the review found insufficient evidence to say whether counter-narratives were effective in preventing violence, but there was some evidence that certain interventions were able to address some of the risk factors relevant to violent extremism (Carthy et al., 2020). Counter-stereotypical exemplars were able to address risk factors including 'realistic perceptions of threat [to one's safety], in-group favouritism and out-group hostility'. However, counter-narratives seem less effective at targeting 'symbolic threat perceptions, implicit bias or intent to act violently' (Carthy et al., 2020, p.3).

As well as the content of counter-narratives, research has sought to understand how to direct those who may be seeking extremist material towards alternative messages. The Redirect method is an example of an online intervention that targets people searching for extremist material aiming to 'prevent unobstructed access to extremist content' (Helmus & Klein, 2018, p.7). First piloted in 2015-16 in 50 states in the USA and funded by Gen Next Foundation, Redirect uses Google's AdWord algorithm to direct individuals seeking extreme-right and militant Islamist material to links that lead to counternarrative videos (Helmus & Klein, 2018; Ganesh, 2019).

An evaluation of Redirect used measures of reach, including impression shares (the percentage of impressions an advert receives in comparison with the total it could achieve), clicks, click-through-rates, and video watch figures to understand its impact (Helmus & Klein, 2018). The evaluation found that the extreme-right campaign received many more impressions than the one targeting militant Islamists, reflecting the greater audience for right-wing extremism in America.

Key Findings

- The effectiveness of counter-narratives varies according to the intervention technique used and the type of outcome targeted.
- There is insufficient evidence to determine whether counter-narratives can prevent violence, however they may be able to address some of the risk factors associated with radicalisation.
- Inoculation theory may provide a foundation for developing deterrence strategies. This approach introduces individuals to weakened versions of an argument whilst providing evidence to refute it. Preliminary experiments indicate that 'active' inoculation methods (where the individual actively engages in a task such as a computer game) can improve critical thinking skills and reduce vulnerability to radicalisation.
- Although the evidence base is very limited, interventions may benefit from adopting a fine-grained approach that is tailored to specific audiences and online contexts, including audience segmentation and micro-targeting.
- Interventions have the potential to produce unintended outcomes, including further entrenching extremist views, for example where activists initiate arguments in response to extremist positions.
- There is some, limited evidence to suggest that highlighting the personal impact of involvement in extremism may be more effective than challenging extremist ideas or arguments, and that online interventions may be less effective with those with more entrenched views.
- Intervention providers working online will benefit from training and support to mitigate the risks associated with this work, and to ensure their approach is evidence informed.

Contemporary Research on CVE Interventions: Counter-narratives

A previous CREST report examining contemporary research on CVE interventions identified a number of findings regarding online counter-narratives (Lewis & Marsden, 2021):

- The existing functionality of mainstream websites such as Facebook and Twitter can be used to reach a potentially large audience of at-risk individuals.
- A more targeted and tailored approach is necessary to successfully engage individuals through alternative platforms and challenge content posted on such sites.
- More work is needed to develop robust methodologies for evaluating the effectiveness of online interventions and counter-narrative campaigns.
- The content of counter-messages and the type of person who delivers them influences their impact. To be effective, the content of messages must resonate with the target audience, and the messenger needs to be seen as credible.
- Participatory methods, which involve target groups in the development and delivery of counter-messages, could be useful in enhancing credibility and resonance.
- Offline intervention providers are increasingly conducting work online. More formal training on how to conduct P/ CVE work online is needed to support this work.

Case Study

COMMUNITY ACTION FOR PREVENTING EXTREMISM, AUSTRALIA

Formerly Exit White Power, Community Action for Preventing Extremism (CAPE) is a community-based initiative, established in 2012, that aims to help young Australians move away from extreme-right groups.

The organisation has developed a range of intervention strategies, including designing counter-narrative material based on monitoring and interpreting far-right output and using ‘Trojan ads’ through Google Adwords to display their content when people searched for extremist material (Brice, 2019). They assessed the impact of this method by the number of click-throughs. The most successful advert was ‘What is white power? The real facts about white power groups in Australia’. The project also went on to be discussed on Stormfront, a far-right Internet forum, which led to increased engagement (Voogt, 2017).

Over time, the project amended its approach to focus less on ideology and more on finding ways to engage directly with those expressing extremist views. Again taking page views as a metric to assess its impact, the project found that material that addressed people’s personal motivations for engaging with the extreme-right were more effective than those that sought to counter ideological claims (Voogt, 2017). This informed their approach to using Facebook advertising to target those who ‘liked’ extremist groups. Their adverts tried to direct users to a ‘White power? Discussion Page’ on Facebook run by the project counsellor which led to 59 discussion threads that attracted thousands of comments (Aly & Lucas, 2015).

A small-scale, preliminary evaluation suggested this approach to using Facebook provided space for users to discuss their views about right-wing extremism, and that ‘the personal identity and social integration function of media should be the foundation for the development of counter-narratives … [p]roviding facts or information that challenges the assumptions of an extremist ideology did not prove to be effective’ (Aly & Lucas, 2015, p.88). Interestingly, the project observed that interventions from anti-racist activists who used the page to challenge the views of people who posted messages, and engage in heated arguments, risked entrenching rather than positively influencing visitors’ attitudes.

Over time, CAPE’s work evolved to focus more on the personal impact of involvement in extremism rather than on debunking ideological claims; increasing civil society and statutory organisations’ capacity to engage in constructive conversations about right-wing extremism; and shifting from an approach that directly questioned people’s views to techniques from counselling including motivational interviewing.

Although unable to assess the effect of the counter-narratives on attitudes or behaviour, the evaluation suggested that Redirect was: 'able to use advertisements linking to counterextremist videos to effectively expose individuals searching for violent jihadist or violent far-right content to content that offered alternative narratives.' (Helmus & Klein, 2018, p. 11). A similar approach using Google Adwords to redirect those seeking right-wing material towards alternative sources of information was used by Exit White Power in Australia, now Community Action for Preventing Extremism (see box on previous page).

Carthy et al.'s review identified a number of risks associated with online interventions. Projects targeting those already convinced by extremist ideas have a lower chance of success because of the common desire to 'maintain psychological consistency' and a reluctance to 'cognitively restructure' existing attitudes (Carthy et al., 2020, p. 30). There is also a risk that using persuasion techniques may produce a 'boomerang effect', further entrenching attitudes rather than reorienting them (Carthy et al., 2020, p. 19). These findings are supported by anecdotal evidence from interviews with practitioners who suggest the online space is 'often a less appropriate tool than offline engagement for dealing with significantly radicalised individuals who are sceptical and less open to opposing viewpoints' (Davey et al., 2019, p.6).

Online interventions can also produce unintended and unanticipated consequences. Unless administered carefully, interventions which, for example, involve mobilising activists against online extremist communities may cause those they are targeting to redouble their commitment to their position (Brice, 2019). Proper training and support for intervention providers is therefore important so they are aware of the risks of different techniques. Support for online intervention providers is also important, as they may not always be alert to the dangers of this work and are often keen to access additional support and training (Davey et al., 2019).

Online interventions are likely to benefit from adopting a fine-grained approach. For example, it is important to understand who the audience is, as well as its size, and who might be observing but not participating in online spaces (Buerger & Wright, 2019). This knowledge can be used to inform the content of the messages, and take account of the identity, perceived influence, and background of those engaged in counter-narrative work (Buerger & Wright, 2019).

Understanding the techniques extremist actors use is also helpful. For instance, some have been described as trying to remove 'grey-zones' to generate 'strategic polarisation', seeking to shape online discourse, and move people towards encrypted channels (Ebner, 2020). Emerging research suggests that focusing on audience segmentation, using micro-targeting, and providing support for those who are targets of online extremists have the potential to contribute to a broad spectrum of online interventions (Ebner, 2020).

4.5.2. 'INOCULATION' TECHNIQUES

An approach attracting increasing attention involves applying attitudinal inoculation theory to countering violent extremism (Braddock, 2020). This involves simultaneously exposing individuals to weakened versions of an argument whilst refuting it (Saleh et al., 2020). Inoculation approaches may be passive, for example through reading text, or active, for instance, using a quiz or game requiring greater cognitive engagement which encourages participants to apply critical thinking skills to the information they encounter.

One of the few empirical studies to analyse the potential for inoculation messages to resist the propaganda and persuasion techniques used in recruitment strategies (n=357) found that preceding extremist content with an inoculation message reduced how persuasive the propaganda was perceived to be (Braddock, 2022). 'Inoculated' participants exhibited greater 'psychological reactance'; perceived the credibility of extremist groups to be lower; and reported lower intentions to support them. Psychological reactance and perceptions of extremist group credibility mediated behavioural intentions to support the group, suggesting that inoculation approaches may be able to counter behavioural radicalisation. As part of the experiment, inoculation messages were attributed to different sources, using left-wing and right-wing propaganda. The study found no evidence that the origin of the message, or the ideological nature of the material moderated the relationships identified.

Inoculation theory was 'gamified' by using games to introduce different kinds of information to participants (n=291) responding to four stages identified in extremist recruitment processes: (1) identification; (2) gaining trust; (3) isolation; and (4) activation (Saleh et al., 2020). A control group played Tetris (n=156) while a treatment group played an online game ('Radicalise') (n=135) that had been constructed for the study's purpose. After quarter of an hour, both groups were presented with six fictional WhatsApp messages designed to evaluate participants' ability to identify manipulation techniques. Each of the messages employed one of the manipulation strategies the treatment group learned whilst playing 'Radicalise'. Preliminary results found that participants who played 'Radicalise' were able to identify manipulative messages with greater confidence and ability, and were more skilled at detecting those factors that make people more vulnerable to recruitment.

4.5.3. PLATFORM SPECIFIC STRATEGIES

Platform specific strategies have also been developed. A study assessing the effectiveness of inoculating participants against Islamophobic and radical-Islamist disinformation (n=591) found similar results to the inoculation studies described above. After watching either a video with material designed to inoculate consumers against extremist messages or a video on an unconnected topic, participants were shown videos including disinformation and 'gateway content that constituted an entry point to potential Islamist or Islamophobic radicalization [sic]' (Lewandowsky & Yesilada, 2021, p.1). Compared with the control group, 'inoculated' participants recorded lower levels of agreement with the messages on the video; reported that they found it less reliable; and were less likely to share it.

However, there are potential risks associated with efforts to make counter-narrative material more accessible. A study using information network analysis assessed two counter-messaging campaigns on YouTube and found that YouTube's automated recommendations system can produce 'relations or endorsements' from counter-messages to extremist content (Schmitt et al., 2018, p. 18). Because of the similarity between the keywords associated with extremist and counter-extremist material, algorithms can generate automatic recommendations that may link these two contrasting types of output.

4.5.4. CONTENT REMOVAL

Motivated by the increased use of the Internet to motivate and, in some cases, live stream terrorist attacks – most notably Brenton Tarrant's 2019 attack on two mosques in New Zealand – law enforcement agencies and tech companies have developed additional methods to counter radicalisation. One response has been the removal of extremist content from online platforms. These efforts are managed at the national level through specific policies and programmes, and at the international level in the context of networks such as the Global Internet Forum to Counter Terrorism. Appropriate methods to interpret the impact of these sorts of initiatives are still being developed, and their outcome is poorly understood (Siegel, 2020; Panday, 2020).

Several studies suggest that this method may be insufficient because removing extremist content does not respond to or challenge these messages (Hassan et al., 2018; Bilazarian, 2020). Other concerns include the implications for freedom of speech (Lowe, 2022; Doucek, 2021); the potential that tools to remove content may be used to shut down opponents, and create a precedent for authoritarian states to follow (Tworek & Leerssen, 2019); the risk that online platforms may 'over-censor in order to avoid the threat of liability' (Douek, 2020, p.41); and the potential that legislation may be outstripping the technical capacity of companies to remove material (Douek, 2020). In addition, because more attention has been paid to militant Islamist than right-wing extremism, algorithms designed to detect extremist content may not pick up right-wing material as easily (Common, 2020). Finally, there are risks to those responsible for content moderation as the extreme nature of the material can produce negative psychological consequences if appropriate strategies are not in place (Reeve, 2021).

4.5.5. CONCLUSION

Empirical evidence assessing intervention and deterrence strategies is not yet robust. Although some research has found evidence that counter-narratives may be effective in addressing some of the risk factors associated with radicalisation, more research is needed to understand when, why and how interventions are effective. Interventions should also take account of the potential unintended consequences of efforts to deter or counter extremism online.

Developing resistance to extremist material using techniques informed by inoculation theory is beginning to develop some empirical support. Future research will benefit from understanding when 'active' strategies aiming to improve participants' abilities to spot manipulative messages, or 'passive' strategies are likely to be more appropriate.

Three areas for further study have been identified (Herath & Whittaker, 2021; Gill et al., 2017; Baugut & Neumann, 2020; Bilarzarian, 2020). These include understanding the potential benefits of combining different tactics to address radicalisation in both online and offline domains, reflecting the growing recognition that radicalisation is a hybrid process informed by experiences across these spaces (Smith et al., 2020).

A better understanding of how to nurture trust between the target audience and those delivering interventions (Bilazarian, 2020), and how to increase audience trust in a counter-narrative's source (Braddock & Morrison, 2020). Finally, researchers have suggested that it will be helpful to replace the relational networks provided by extremist groups (found both online and offline) by cultivating strategies which offer alternative sources of meaning and connection (Bilazarian, 2020).

4.6.CHALLENGES TO UNDERSTANDING ONLINE RADICALISATION

This section provides a brief outline of the barriers and challenges to understanding the role of the Internet in radicalisation processes. These include difficulties accessing data that is able to determine whether online material shapes attitudes and behaviours; challenges facing efforts to interpret changes in user sentiment towards extremist material; and the difficulty understanding who, of all those who might encounter online extremist content, might find it most persuasive.

Methodological challenges also limit the conclusions it is currently possible to draw, both regarding whether and how engaging with extremist material might shape the move towards violent extremism, and what works to divert people away from adopting extremist attitudes. Disciplinary and conceptual differences in how radicalisation and the role of the Internet are understood can also make it difficult to develop a robust evidence base.

4.6.1. ACCESSING DATA

It is difficult to measure or collect data on Internet users' opinions about extremist content, and the role online material plays in shaping behavioural or cognitive outcomes. Studies which attempt to assess the extent of user behaviour and interaction with online content tend to quantify the number of likes or dislikes posts receive, or how many times content is shared (Bilazarian, 2020). However, this does not capture user sentiment or whether, and to what extent, user opinion shifts following their engagement with online material.

As Section 4.4 describes, it has also proven difficult to understand the relationship between online engagement with extremist content and settings and offline violence. Tracing individual pathways across online and offline spaces is challenging, ethically,

Key Findings

- Accessing and gathering valid empirical data is one of the main barriers to producing robust research able to evidence whether, and to what extent, online activity influences violent offline behaviour. Similar difficulties arise in efforts to assess what influences attitudinal change.
- It can be difficult to generalise the findings of the research drawn from small-n sample sizes collected using qualitative methods, or which focus on a specific ideology or geographical context. Drawing broader conclusions to groups or settings beyond the data sample should be undertaken with caution.
- Large-n computational methods have the potential to identify broader trends in the data but can risk over-simplifying radicalisation processes.
- Efforts to understand the impact of online interventions face similar challenges to evaluations of offline P/CVE programmes. These include the difficulty understanding an intervention's impact; accessing appropriate data; ethical and security risks; and the difficulty identifying and evidencing the causal factors that shape outcomes.
- Methodological differences in how data are collected, used and analysed can be difficult to translate across disciplines.
- Ambiguous and/ or contested definitions of 'online radicalisation' can make it challenging to draw comparisons across studies which may be focused on different phenomena.

analytically, and practically. Furthermore, it is difficult to determine what role the online space plays without a more contextualised appreciation and understanding of individual trajectories. Given the range of influences an individual may be subject to, and the hybrid nature of radicalisation pathways, understanding the role of online extremist content raises particular challenges.

4.6.2. METHODOLOGICAL CHALLENGES

Methodologically, research tends to use large-scale quantitative approaches or small-n qualitative methods. Both have benefits, however, the relatively small size of the samples used in many qualitative studies mean they are not easily generalisable beyond the specific sample. Such studies also typically have a relatively limited ideological or geographical focus, which although they can produce nuanced findings, are less able to support broader conclusions about online radicalisation.

Large-n computational methods can provide broader analyses of trends in the data. However, data mining large sets of data - common in studies that 'scrape' data from social media platforms - and use of natural language processing techniques have been critiqued for over-simplifying radicalisation processes and being unable to account for the individualised nature of pathways into and out of extremism (Ajala et al., 2022; El Barachi et al., 2022). There is also the risk that methodologies that focus exclusively on online processes may neglect offline influences and the hybrid nature of radicalisation processes.

Research into the effectiveness of interventions and deterrence strategies suffers from many of the challenges facing efforts to interpret the impact of offline P/CVE programmes. These include the difficulty understanding the specific impact of an intervention, given the range of other influences that may be at work; the challenges accessing appropriate data; ethical and security risks associated with testing online interventions; and the difficulty identifying and evidencing the causal factors that might make the appropriate target of interventions, and through which it might be possible to assess their effects (see Lewis et al., 2020).

4.6.3. DEFINITIONAL AMBIGUITY AND INTERDISCIPLINARY TENSIONS

Online radicalisation research is carried out in a range of different disciplines, and while this can be a strength of the literature, it can also create challenges. The use of a wide range of study designs and methodological perspectives offers a variety of ways of understanding the topic. Such diversity contributes to the development of innovative methods and research designs.

However, there is something of a disjuncture between research conducted in the social sciences and that in cyber security studies and computer sciences. Differences in how data are collected, used and analysed can make it difficult to translate findings across disciplines and there is a risk that research is siloed such that researchers are unaware of work going on in other fields. There is also some ambiguity as to how 'online radicalisation' is defined. Studies use different variables and conceptualise the process of online radicalisation in different ways which can make comparing findings difficult.

4.6.4. CONCLUSION

Research into online radicalisation faces several barriers and challenges, some of which are extremely difficult to address. There is much more work to do to identify ways of deriving and accessing robust empirical data able to show user sentiment towards extremist content or which captures evidence of whether, how, and the extent to which, online activity influences violent offline behaviour. Challenges surrounding definitional ambiguity and interdisciplinary tensions will benefit from greater collaboration between researchers working across different disciplines and more opportunities to share research between relevant fields.

5. CONCLUSIONS

5.1. KEY FINDINGS

- Online and offline activities and domains interact, challenging the 'online/offline dichotomy' popular in early research into online radicalisation. Radicalisation processes rarely take place in either the online domain or the offline sphere exclusively, but instead are characterised by complex and dynamic interactions between the two.

- Research that sought to distinguish between online and offline processes may have over-estimated the extent to which the Internet contributes to radicalisation processes. This tendency to focus on the role of the Internet may have come at the expense of recognising the role of offline factors.

- The Internet in isolation does not cause radicalisation and is better understood as facilitating this process. While the Internet can contribute to an individual's radicalisation, it cannot drive the process on its own.

BEHAVIOURAL RADICALISATION

- Use of the Internet can enable behavioural outcomes including event planning and preparatory activities, communication and networking behaviours (including arranging offline activities) and ideology-seeking actions.

- Pathways into violent extremism have been characterised as primarily offline, mainly online, and hybrid. Hybrid pathways seem to be the most common.

- There is no single profile of, or standard trajectory taken by, individuals whose use of the Internet influenced their radicalisation. However different pathways seem to be associated with differing levels of intent, capability, and engagement. Hybrid pathways demonstrate greatest engagement and intent; offline pathways, greatest capability; and online, the lowest levels of engagement, intent and capability.

COGNITIVE RADICALISATION

- Empirical research analysing the influence of online interactions and exposure to extremist content on violent extremist behaviour remains limited.

- Video-sharing platforms and social networking sites are spaces where individuals are most likely to encounter extremist content online.

- The individual is an active rather than passive actor in the radicalisation process. It is the individual's behaviour and how they utilise the Internet that informs its relevance to radicalisation.

- There is little robust evidence about whether and how recruiters try to identify or engage with those seeking out online extremist material.

- Individuals who actively seek out violent extremist material online seem to be at greater risk of radicalising and engaging in violence, compared to passive consumers.

- Research on the role exposure to violent extremist content online plays in cognitive radicalisation has suggested that initial exposure to extremist content online has the potential to trigger an interest in extreme ideologies, and that exposure to content from a combination of online and offline spheres may be more influential than exposure via one or the other.

- The amount of time spent online and willingness to express political views on the Internet seem to be associated with greater exposure to extremist material.

- Personality traits, specifically aggression, may be more influential than exposure to extremist propaganda in influencing extremist cognitions.

ONLINE INDICATORS OF BEHAVIOURAL RADICALISATION

- Robust empirical evidence on how online activities might be used to identify individuals at risk of behavioural radicalisation is comparatively weak.
- There is some evidence that exposure to online extremist content has a stronger link to radicalisation in comparison with other kinds of media-related risk factors.
- Recruiters may use different kinds of online extremist material to first nurture cognitive radicalisation and then try and move people towards violence.
- Some research suggests that posting patterns on social media may be able to differentiate between violent and non-violent extremists, and between behavioural and cognitive outcomes, but further research is needed to fully understand these processes.
- Future research is likely to benefit from combining computational and social science methods, and developing robust, publicly available standardised datasets which are free from bias.

INTERVENTION STRATEGIES

- The effectiveness of counter-narratives varies according to the intervention technique used and the type of outcome targeted.
- There is insufficient evidence to determine whether counter-narratives can prevent violence, however they may be able to address some of the risk factors associated with radicalisation.
- Inoculation theory may provide a foundation for developing deterrence strategies. This approach introduces individuals to weakened versions of an argument whilst providing evidence to refute it. Preliminary experiments indicate that ‘active’ inoculation methods (where the individual actively engages in a task such as a computer game) can improve critical thinking skills and reduce vulnerability to radicalisation.
- Although the evidence base is very limited, interventions may benefit from adopting a fine-grained approach that is tailored to specific audiences and online contexts, including audience segmentation and micro-targeting.
- Interventions have the potential to produce unintended outcomes, including further entrenching extremist views, for example where activists initiate arguments in response to extremist positions.
- There is some, limited evidence to suggest that highlighting the personal impact of involvement in extremism may be more effective than challenging extremist ideas or arguments, and that online interventions may be less effective with those with more entrenched views.
- Intervention providers working online will benefit from training and support to mitigate the risks associated with this work, and to ensure their approach is evidence-informed.

CHALLENGES TO UNDERSTANDING ONLINE RADICALISATION

- Accessing and gathering valid empirical data is one of the main barriers to producing robust research able to evidence whether, and to what extent, online activity influences violent offline behaviour. Similar difficulties arise in efforts to assess which factors influence attitudinal change.
- It can be difficult to generalise the findings the research drawn from small-n sample sizes collected using qualitative methods, or which

focuses on a specific ideology or geographical context. Drawing broader conclusions to groups or settings beyond the data sample should be undertaken with caution.

- Large-n computational methods have the potential to identify broader trends in the data but can risk over-simplifying radicalisation processes.
- Efforts to understand the impact of online interventions face similar challenges to evaluations of offline P/CVE programmes. These include the difficulty understanding an intervention's impact; accessing appropriate data; ethical and security risks; and the difficulty identifying and evidencing the causal factors that shape outcomes.
- Methodological differences in how data are collected, used and analysed can be difficult to translate across disciplines.
- Ambiguous and/ or contested definitions of 'online radicalisation' can make it challenging to draw comparisons across studies which may be focused on different phenomena.

5.2. RECOMMENDATIONS FOR POLICY AND PRACTICE

- P/CVE interventions are likely to benefit from taking account of the hybrid nature of radicalisation processes and developing ways of targeting online and offline domains simultaneously, rather than separately.
- Intervention strategies which provide an alternative source of meaning and association to replace the relational networks offered by extremist groups, both online and offline, appear promising.
- There is some evidence to suggest it may be beneficial to prioritise interventions which focus on those who actively seek extremist content online, as they may be at greater risk of radicalisation to violence.
- The gamification (or use of mechanisms used in games) of interventions has the potential to appeal to those who actively seek extremist content. These types of intervention can encourage the development of critical thinking skills and may provide an element of interaction that active seekers are looking for.
- Interventions targeting video-sharing platforms and social networking sites may have a greater impact than targeting other areas online. However, there are risks to this approach. Counter-messaging videos and extremist content can share key words. This means that the algorithms which drive automated recommendation systems may direct users to extremist content, rather than to counter-messaging videos.
- Counter-narratives will benefit from careful targeting, taking account of the specific audience; the extent to which they may already be persuaded by extremist ideas; the risk factors the intervention is seeking to influence and the mechanisms by which positive outcomes might be enabled.
- Evidence regarding the impact of removing extremist content is limited. Although taking down material may help to reduce its accessibility, it carries risks including the potential to encourage users to move to encrypted platforms which are more difficult to monitor and moderate.
- Interventions should take account of unintended outcomes, including the potential to further entrench extremist views; generate risks to freedom of speech; and create incentives for tech companies to 'over-censor' content to avoid sanction.
- Intervention providers working online should receive appropriate training, professional development opportunities, and support.

5.3. DIRECTIONS FOR FUTURE RESEARCH

KEY AREAS OF FUTURE RESEARCH INCLUDE:

- Further work to understand the role of the Internet in pathways into extremism, including research able to interpret how online and offline dynamics interact.

- Research that draws on first-hand accounts of how the Internet shaped an individual's thinking and behaviour has the potential to elucidate the experiential aspects of radicalisation processes.

- Studies examining the impact of the COVID-19 pandemic on online radicalisation could try to assess the impact of lockdowns and whether associated feelings of isolation and the increased use of technology as a substitute for physical, face-to-face interactions led to greater exposure to, or engagement with, extremist content.

- Research which bridges computational approaches which analyse large amounts of data with social science-based methods able to interpret the experiential and subjective experiences of online users may provide greater insights and overcome the disjuncture between disciplines.

- Studies focused on a specific ideology could be carried out with data on a different ideology. This would help to determine whether findings can be generalised or are ideologically specific, and whether targeted interventions would benefit from being tailored to specific ideologies.

- Further research into the role of individual personality traits, pre-existing beliefs and other psychological factors that may shape responses to extremist content and radicalisation. This would help tailor and target interventions in ways which are appropriate for particular groups or individuals, and help to avoid unintended or negative outcomes.

- Areas where results are limited, mixed or inconclusive would benefit from further research. These include:

  a. The relationship between exposure to extremist content online and cognitive radicalisation.

  b. Approaches able to interpret whether patterns of online engagement have the potential to identify individuals at risk of cognitive or behavioural radicalisation.

- Further work to understand the impact of interventions is important, assessing:

  a. What effect the removal of online extremist content has, and what risks this strategy carries.

  b. The potential of realist evaluation to develop a better understanding of which counter-narrative interventions work, for whom, under what circumstances, and why.

  c. The unintended consequences of different kinds of intervention strategy, including direct engagement online; efforts to direct people to counter messages; and counter-narrative material.

BIBLIOGRAPHY

Ajala, I., Feroze, S., El Barachi, M., Oroumchian, F., Mathew, S., Yasin, R., & Lutfi, S. (2022). Combining artificial intelligence and expert content analysis to explore radical views on twitter: Case study on far-right discourse. *Journal of Cleaner Production, 362,* 132263, 1-17. *https://doi.org/10.1016/j.jclepro.2022.132263*

Aly, A., & Lucas, K. (2015). 'Countering Online Violent Extremism in Australia: Research and Preliminary Findings'. In S. Zeiger & A. Aly (eds.) *Countering Violent Extremism: Developing an Evidence-Base for Policy and Practice*. Perth, WA: Hedayah and Curtin University, pp. 81-89.

Araque, O. & Iglesias, C. A., (2022). An Ensemble Method for Radicalization and Hate Speech Detection Online Empowered by Sentic Computing. *Cognitive Computation, 14*, 48-61. *https://doi.org/10.1007/s12559-021-09845-6*

Badawy, A. & Ferrara, E. (2018). The rise of Jihadist propaganda on social networks. *Journal of Computational Social Science, 1*, 453-470. *https://doi.org/10.1007/s42001-018-0015-z*

Bastug, M. F., Douai, A., & Akca, D. (2020). Exploring the "Demand Side" of Online Radicalization: Evidence from the Canadian Context. *Studies in Conflict and Terrorism, 43*(7), 616–637. *https://doi.org/10.1080/1057610X.2018.1494409*

Baugut, P., & Neumann, K. (2020). Online propaganda use during Islamist radicalization. *Information Communication and Society*, *23*(11), 1570–1592. *https://doi.org/10.1080/1369118X.2019.1594333*

Buerger, C. & Wright, L. (2019). 'Counterspeech: A literature review'. *Dangerous Speech*. Available at: *https://dangerousspeech.org/wp-content/uploads/2019/11/Counterspeech-lit-review_complete-11.20.19-2.pdf*

Bilazarian, T. (2020). Countering Violent Extremist Narratives Online: Lessons From Offline Countering Violent Extremism. *Policy and Internet*, *12*(1), 46–65. *https://doi.org/10.1002/poi3.204*

Braddock, K. (2022). Vaccinating Against Hate: Using Attitudinal Inoculation to Confer Resistance to Persuasion by Extremist Propaganda. *Terrorism and Political Violence*, *34*(2), 240–262. *https://doi.org/10.1080/09546553.2019.1693370*

Braddock, K. (2020). *Weaponized words: The strategic role of persuasion in violent radicalization and counter-radicalization.* Cambridge, Cambridge University Press. *https://doi.org/10.1017/9781108584517*

Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict & Terrorism, 39*(5), 381-404. *https://doi.org/10.1080/1057610X.2015.1116277*

Braddock, K., & Morrison, J. F. (2020). Cultivating Trust and Perceptions of Source Credibility in Online Counternarratives Intended to Reduce Support for Terrorism. *Studies in Conflict & Terrorism, 43*(6), 468-492. *https://doi.org/10.1080/1057610X.2018.1452728*

Brice, P. (2019). 'Challenging the Far-Right in Australia'. In M. Peucker & D. Smith (eds.), *The Far-Right in Contemporary Australia*. Singapore: Palgrave Macmillan, pp.199-214. *https://doi.org/10.1007/978-981-13-8351-9*

Carthy, S. L., Doody, C. B., Cox, K., O'Hora, D., & Sarma, K. M. (2020). Counter-narratives for the prevention of violent radicalisation: A systematic review of targeted interventions. *Campbell Systematic Reviews*, *16*(3), 1-37. *https://doi.org/10.1002/cl2.1106*

Common, M. F. (2020). Fear the Reaper: how content moderation rules are enforced on social media. *International Review of Law, Computers & Technology, 34*(2), 126-152. *https://doi.org/10.1080/13600869.2020.1733762*

Conway, M. (2017). Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Studies in Conflict and Terrorism*, *40*(1), 77–98. *https://doi.org/10.1080/1057610X.2016.1157408*

Costello, M., Barrett-Fox, R., Bernatzky, C., Hawdon, J. & Mendes, K. (2020). Predictors of Viewing Online Extremism Among America's Youth. *Youth & Society*, *52*(2), 710-727. https:// *doi.org/10.1177/0044118X18768115*

Davey, J., Tuck, H. and Amarasingam, A. (2019). *An imprecise science: Assessing interventions for the prevention, disengagement and de-radicalisation of left and right-wing extremists*. Institute for Strategic Dialogue. Available at: *https://www.isdglobal.org/isd-publications/an-imprecise-science-assessing-interventions-for-the-prevention-disengagement-and-de-radicalisation-of-left-and-right-wing-extremists/*

Douek, E. (2020). Australia's 'Abhorrent Violent Material' Law: Shouting 'Nerd Harder' and Drowning Out Speech. *Australian Law Journal 94*, 41-60, Available at: *https://ssrn.com/abstract=3443220*

Douek, E. (2021). The Limits of International Law in Content Moderation. *University of California, Irvine (UCI) Journal of International, Transnational, and Comparative Law* 6(1), 37-76. Available at: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3709566*

Ebner, J. (2020). *Going Dark: The Secret Social Lives of Extremists*. London; Dublin, Bloomsbury Publishing.

El Barachi, M., Mathew, S. S., Oroumchian, F., Ajala, I., Lutfi, S., & Yasin, R. (2022). Leveraging Natural Language Processing to Analyse the Temporal Behavior of Extremists on Social Media. *Journal of Communications Software and Systems, 18*(2), 195-207. *https://doi.org/10.24138/jcomss-2022-0031*

Fernandez, M., Gonzalez-Pardo, A., & Alani, H. (2019). Radicalisation Influence in Social Media. *Journal of Web Science, 6*, 1–15. *https://doi.org/10.1561/106*

Fernandez, M., & Harith, A. (2021). Artificial Intelligence and Online Extremism: Challenges and Opportunities. In J. McDaniel & K. Pease (eds.), *Predictive Policing and Artificial Intelligence.* Abingdon; New York: Routledge, pp. 132-162. *https://doi.org/10.4324/9780429265365-7*

Frissen, T. (2021). Internet, the great radicalizer? Exploring relationships between seeking for online extremist materials and cognitive radicalization in young adults. *Computers in Human Behavior*, *114*(August 2020), 106549. *https://doi.org/10.1016/j.chb.2020.106549*

Frissen, T., & d'Haenens, L. (2017). Legitimizing the caliphate and its politics: Moral disengagement rhetoric in Dabiq. In S.F. Krishna-Hensel (ed.), *Authoritarian and Populist Influences in the New Media*. Abingdon: Routledge. *https://doi.org/10.4324/9781315162744*

Frissen, T., Toguslu, E., Van Ostaeyen, P., & d'Haenens, L. (2018). Capitalizing on the koran to fuel online violent radicalization: A taxonomy of Koranic references in ISIS's Dabiq. *Telematics and Informatics, 35*(2), 491–503. *https://doi.org/10.1016/j.tele.2018.01.008*

Gaikwad, M., Ahirrao, S., Phansalkar, S., & Kotecha, K. (2021). Online Extremism Detection: A Systematic Literature Review with Emphasis on Datasets, Classification Techniques, Validation Methods, and Tools. *IEEE Access*, *9*, 48364–48404. *https://doi.org/10.1109/ACCESS.2021.3068313*

Ganesh, B. (2019). Evaluating the Promise of Formal Counter-Narratives. In B. Ganesh & J. Bright (eds.), *Extreme digital speech: Contexts, responses and solutions*. VOX-Pol, pp. 89-98. Available at: *https://www.voxpol.eu/download/vox-pol_publication/DCUJ770-VOX-Extreme-Digital-Speech.pdf#page=91*

Gaudette, T., Scrivens, R., & Venkatesh, V. (2020). The Role of the Internet in Facilitating Violent Extremism: Insights from Former Right-Wing Extremists. *Terrorism and Political Violence*, *00*(00), 1–18. *https://doi.org/10.1080/09546553.2020.1784147*

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes. *Criminology and Public Policy*, *16*(1), 99–117. *https://doi.org/10.1111/1745-9133.12249*

Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., Rousseau, C., & Sieckelinck, S. (2018). Exposure to Extremist Online Content Could Lead to Violent Radicalization: A Systematic Review of Empirical Evidence. *International Journal of Developmental Sciences*, *12*(1–2), 71–88. *https://doi.org/10.3233/DEV-170233*

Hawdon, J., Bernatzky, C. & Costello, M. (2019). Exposure to Violent Extremist Material Online: Cyber-Routines, Political Attitudes, and Exposure to Violence-Advocating Online Extremism. *Social Forces*, *98*(1), 329-354. *https://doi.org/10.1093/sf/soy115*

Helmus, T. C. & Klein, K. (2018). Assessing Outcomes of Online Campaigns Countering Violent Extremism: A Case Study of the Redirect Method. Santa Monica, CA: RAND Corporation. Available at: *https://www.rand.org/pubs/research_reports/RR2813.html*.

Herath, C., & Whittaker, J. (2021). Online Radicalisation: Moving beyond a Simple Dichotomy. *Terrorism and Political Violence*, *00*(00), 1–22. *https://doi.org/10.1080/09546553.2021.1998008*

HM Government (2012). Channel: Vulnerability assessment framework. October. London: United Kingdom. Available at *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/118187/vul-assessment.pdf*

Howard, T., Poston, B., & Lopez, A. (2022). Extremist Radicalization in the Virtual Era: Analyzing the Neurocognitive Process of Online Radicalization. *Studies in Conflict and Terrorism*, *0*(0), 1–26. *https://doi.org/10.1080/1057610X.2021.2016558*

Institute for Economics & Peace (2022). Global Terrorism Index 2022: Measuring the Impact of Terrorism. March. Sydney: Australia. Available at *https://www.visionofhumanity.org/resources/*.

Jensen, M. A., Atwell Seate, A., & James, P. A. (2020). Radicalization to violence: A pathway approach to studying extremism. *Terrorism and Political Violence, 32*(5), 1067–1090. *https://doi.org/10.1080/09546553* .2018.1442330

Kenyon, J., Binder, J. F., & Baker-Beall, C. (2022). Online radicalization: Profile and risk analysis of individuals convicted of extremist offences. *Legal and Criminological Psychology*, *April*, 1–17. *https://doi.org/10.1111/lcrp.12218*

Koehler, D. (2014). The radical online: Individual radicalization processes and the role of the Internet. *Journal for Deradicalization*, *1* (Winter), 116–134. Available at: *https://journals.sfu.ca/jd/index.php/jd/article/view/8http*://*journals.sfu.ca/jd/index.php/jd/article/view/8%5Cnhttp*://*journals.sfu.ca/jd/index.php/jd/article/download/8/8*

Lara-Cabrera, R., Gonzalez-Perez, A., Benouaret, K., Faci, N., Benslimane, D., & Camacho, D. (2017). Measuring the Radicalisation Risk in Social Networks. *IEEEAccess, 5*, 10892-10900. *https://doi.org/10.1109/ACCESS.2017.2706018*

Lara-Cabrera, R., Gonzalez-Perez, A., & Camacho, D. (2019). Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in Twitter. *Future Generation Computer Systems, 93*, 971-978. *https://doi.org/10.1016/j.future.2017.10.046*

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

Lewis, J., Marsden, S. & Copeland, S. (2020). Evaluating Programmes To Prevent And Counter Extremism. Lancaster University, Lancaster: Centre for Research and Evidence on Security Threats (CREST). Available at: *https://crestresearch.ac.uk/resources/evaluating-programmes-to-prevent-and-counter-extremism*

Lewis, J. & Marsden, S. V. (2021). Countering Violent Extremism Interventions: Contemporary Research. Lancaster: Centre for Research and Evidence on Security Threats (CREST). Available at: *https://crestresearch.ac.uk/resources/countering-violent-extremism-interventions/*

Lowe, D. (2022). Far-Right Extremism: Is it Legitimate Freedom of Expression, Hate Crime, or Terrorism?. *Terrorism and Political Violence*, *34*(7), 1433-1453. *https://doi.org/10.1080/09546553.2020.1789111*

Macdonald, S., & J. Whittaker, J., (2019) “Online Radicalization: Contested Terms and Conceptual Clarity,” In J.R. Vacca (ed.), *Online Terrorist Propaganda, Recruitment, and Radicalisation.* Boca Raton: CRC Press, pp. 33-46. *https://doi.org/10.1201/9781315170251*

Meleagrou-Hitchens, A., Alexander, A., & Kaderbhai, N. (2017). The impact of digital communications technology on radicalization and recruitment. *International Affairs*, *93*(5), 1233–1249. *https://doi.org/10.1093/ia/iix103*

Mills, C. E., Freilich, J. D., Chermak, S. M., Holt, T. J., & LaFree, G. (2021). Social Learning and Social Control in the Off- and Online Pathways to Hate Crime and Terrorist Violence. *Studies in Conflict and Terrorism*, *44*(9), 701–729. *https://doi.org/10.1080/1057610X.2019.1585628*

Nagy, Z. (2018). *Artificial Intelligence and Machine Learning Fundamentals: Develop Real-World Applications Powered by the Latest AI Advances,* Birmingham, Packt Publishing, Limited.

Nienierza, A., Reinemann, C., Fawzi, N., Riesmeyer, C. & Neumann, K. (2021). Too dark to see? Explaining adolescents’ contact with online extremism and their ability to recognize it. *Information, Communication & Society*, *24*(9), 1229-1246. *https://doi.org/10.1080/1369118X.2019.1697339*

Odağ, Ö., Leiser, A., & Boehnke, K. (2019). Reviewing the role of the internet in radicalization processes. *Journal for Deradicalization, 21* (Winter), 261–300. Available at: *https://journals.sfu.ca/jd/index.php/jd/article/view/289*

Panday, P. (2020). 'One year since the Christchurch Call to Action: A Review'. *ORF Issue Brief No. 389,* August 2020. New Delhi, India: Observer Research Foundation. Available at: *https://www.orfonline.org/research/one-year-since-the-christchurch-call-to-action-a-review/*

Pauwels, L., & Schils, N. (2016). Differential online exposure to extremist content and political violence: Testing the relative strength of social learning and competing perspectives. *Terrorism and Political Violence, 28,* 1-29. *https://doi.org/10.1080/09546553.2013.876414*

Reeve, Z. (2021). Engaging with Online Extremist Material: Experimental Evidence. *Terrorism and Political Violence*, *33*(8), 1595–1620. *https://doi.org/10.1080/09546553.2019.1634559*

Roiger, R. (2017). *Data Mining: A Tutorial Based Primer.* 2nd edition. Boca Raton: CRC Press. *https://doi.org/10.1201/9781315382586*

Saleh, N. F., Roozenbeek, J., Makki, F. A., McClanahan, W. P., & van der Linden, S. (2021). Active inoculation boosts attitudinal resistance against extremist persuasion techniques: a novel approach towards the prevention of violent extremism. *Behavioural Public Policy*, 1–24. *https://doi.org/10.1017/bpp.2020.60*

Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! the potential role of YouTube. *Journal of Communication*, *68*(4), 758–779. *https://doi.org/10.1093/joc/jqy029*

Scrivens, R., Gill, P., & Conway, M. (2020). The Role of the Internet in Facilitating Violent Extremism and Terrorism: Suggestions for Progressing Research. In T. Holt & A. Bossler (eds.) *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Cham: Palgrave Macmillan, pp. 1417–1435. *https://doi.org/10.1007/978-3-319-78440-3_61*

Scrivens, R., Wojciechowski, T. W., Freilich, J. D., Chermak, S. M., & Frank, R. (2021). Comparing the Online Posting Behaviors of Violent and Non-Violent Right-Wing Extremists. *Terrorism and Political Violence*, *00*(00), 1–18. *https://doi.org/10.1080/09546553.2021.1891893*

Siegel, A. A. (2020). 'Online hate speech'. *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge: Cambridge University Press, pp. 56-88. *https://doi.org/10.1017/9781108890960*

Shortland, N., & McGarry, P. (2022). Supplemental Material for The Personality and Propaganda Puzzle: Exploring the Effect of Personality on Exposure to Extremist Content Online. *Psychology of Violence*, *12*(1), 1–10. *https://doi.org/10.1037/vio0000396.supp*

Shortland, N., Nader, E., Thompson, L. & Palasinski, M. (2022). Is Extreme in the Eye of the Beholder? An Experimental Assessment of Extremist Cognitions. *Journal of Interpersonal Violence*, *37*(7-8), NP4865-NP4888. *https://doi.org/10.1177/0886260520958645*

Smith, L. G. E., Blackwood, L., & Thomas, E. F. (2020). The Need to Refocus on the Group as the Site of Radicalization. *Perspectives on Psychological Science*, *15*(2), 327–352. *https://doi.org/10.1177/1745691619885870*

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics, 17*(1), 168-192, *https://doi.org/10.1016/j.aci.2018.08.003*

Theodosiadou, O., Pantelidou, K., Bastas, N., Chatzakou, D., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. (2021). Change point detection in terrorism-related online content using deep learning derived indicators. *Information (Switzerland)*, *12*(7). *https://doi.org/10.3390/info12070274*

Tworek, H. & Leerssen, P. (2019). *An analysis of Germany's NetzDG law*. Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. Available at: *https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf*

Valentini, D., Lorusso, A. M., & Stephan, A. (2020). Onlife Extremism: Dynamic Integration of Digital and Physical Spaces in Radicalization. *Frontiers in Psychology*, *11*(March). *https://doi.org/10.3389/fpsyg.2020.00524*

von Behr, I., Reding, A., Edwards, C., & Gribbon, L. (2013). *Radicalisation in the digital era: The use of the internet in 15 cases of terrorism and extremism.* Santa Monica, CA: RAND Corporation. *https://doi.org/10.7249/RR453*

Voogt, S. (2017). Countering far-right recruitment online: CAPE's practitioner experience. *Journal of Policing, Intelligence and Counter Terrorism, 12*(1), 34-46. *https://doi.org/10.1080/18335330.2016.1215510*

Wendelberg, L. (2021). An Ontological Framework to Facilitate Early Detection of 'Radicalization' (OFEDR)—A Three World Perspective. *Journal of Imaging*, *7*(60), 1-27, *https://doi.org/10.3390/jimaging* 7030060

Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, J. (2020). Online extremism: Research trends in internet activism, radicalization, and counter-strategies. *International Journal of Conflict and Violence*, *14*(2), 1–20. *https://doi.org/10.4119/ijcv-3809*

Wolfowicz, M., Perry, S., Hasisi, B., & Weisburd, D. (2021). Faces of radicalism: Differentiating between violent and non-violent radicals by their social media profiles. *Computers in Human Behavior*, *116* (November 2020), 106646, 1-10. *https://doi.org/10.1016/j.chb.2020.106646*

Wolfowicz, M., Hasis, B. & Weisburd, D. (2022). What are the effects of different elements of media on radicalization outcomes? A systematic review. *Campbell Systematic Reviews, 18,* e1244, 1-50. *https://doi.org/10.1002/cl2.1244*

Youngblood, M. (2020). Extremist ideology as a complex contagion: the spread of far-right radicalization in the United States between 2005 and 2017. *Humanities & Social Sciences Communications, 7*(49), 1-10. *https://doi.org/10.1057/s41599-020-00546-3*

For more information on CREST
and other CREST resources, visit
www.crestresearch.ac.uk

CREST

CENTRE FOR RESEARCH AND
EVIDENCE ON SECURITY THREATS