

Review

Public Mental Health Approaches to Online Radicalisation: a Systematic Review of the Literature

Rabya Mughal ^{1,*}, Valerie DeMarinis ^{2,3}, Maria Nordendahl ², Hassan Lone⁴, Veronica Phillips⁵ and Eolene Boyd-MacMillan¹

¹ Department of Psychiatry, Cambridge Public Health, University of Cambridge, United Kingdom.

² Department of Public Health and Clinical Medicine, Umeå University, Sweden.

³ Innlandet Hospital Trust, Norway.

⁴ St George's Medical School, University of London, United Kingdom.

⁵ University of Cambridge Medical Library, United Kingdom

* Correspondence: rm667@cam.ac.uk

Abstract: This systematic review seeks to position online radicalisation within whole system frameworks incorporating individual, family, community and wider structural influences, whilst reporting evidence of public mental health approaches for individuals engaging in radical online content. Methods: Authors searched Medline (via Ovid), PsycInfo (via Ebscohost) and Web of Science (Core Collection), with the use of Boolean operators across 'extremism', 'online content', and 'intervention'. Results: Following assessment of full text, all retrieved papers had to be excluded. Results from six excluded articles which did not fit inclusion criteria but identified theoretical relationships between all three elements of online extremism, psychological outcomes, and intervention strategy, were discussed. Authors found no articles outlining public mental health approaches to specifically online radicalisation. Conclusions: There is an immediate need for further research in this field given the increase in different factions of radicalised beliefs resulting from online, particularly social media, usage.

Keywords: radicalisation; social media; public mental health; online radicalisation

1. Introduction

The existence and efficacy of a public mental health approach to online radicalisation is not clear. Notwithstanding debates regarding whether there is a relationship between psychological health and radicalised beliefs [1,2], and concerns with the use of mental health services in counter-terrorism policy [3-5], the number of initiatives within this field continues to grow. These include risk assessment protocols for vulnerable individuals [3], mental health referrals within the UK [6], and counselling and family therapy approaches in de-radicalisation [7]. However, whether such approaches are effective in the long term is not currently known [4,5]. Despite goals that can overlap with capacity building and community empowering initiatives, public mental health promotion, as part of a general public health framework, exists outside of counter-terrorism and incorporates whole-system frameworks where individual, family, community and wider structural influences contribute to subjective wellbeing [8]. Wellbeing on an individual level for instance can be influenced by sociodemographic factors (such as income, housing and employment), physical and psychological health, resilience, identity factors, adverse childhood experiences and/or trauma, amongst a number of other risk/ protective factors [8]. On a structural level, wellbeing can be influenced by societal discrimination, social and cultural norms, economic conditions, inequality, political structures, global politics and/or migration, amongst a number of other risk/ protective factors [8]. Radicalisation depends on complex interactions between different risk factors - a disruption in one or several of these can lead to a multitude of negative psychosocial outcomes, one of which might be dissatisfaction with authority, a need to belong, a propensity to favour populist policy, or joining an extreme organisation [9].

One purpose of this paper was to understand the positioning of specifically *online* radicalisation within such whole-system frameworks. The need has grown since the landscape of radicalisation has become increasingly nebulous, now including (but not limited to) far-right, far-left, conspiracy theory, 'incel' and 'eco-activism', with several populist agendas appearing within mainstream politics, along with the fluidity of online-offline spaces [10-13]. The vast majority of internet users adopt a number of media types, including video hosting, social media, peer to peer, gaming and livestreaming platforms. For example, on TikTok (a collaborative, short form video hosting service):

“After watching a thousand videos [around seventeen viewing hours], recommendations [by the platform to users] become increasingly radical in nature, in content, and in tone, ultimately sending [an individual] into conspiratorial echo chambers ... Mainstream conservative political material [turns] toward antivaxxer material, hypermasculinity ... Hatescape, a conspiracy rabbit hole, socialization and education on hate or dissidence, and even some calls for and demonstration of violence” [14].

A number of social media, video hosting and peer to peer networks have been subject to criticism due to their contribution to psychological effects such as addiction, attention deficiency, individual and group attitudinal shift, as well as controversies regarding inappropriate content, misinformation, disinformation, moderation, user privacy and censorship [15].

The aim of this review was to find public mental health approaches that have been utilised to identify, prevent, or address online ideological and political radicalisation. In addition, this review sought to identify the nature of online content, the demographics of those accessing content, and wider correlates between radicalised beliefs and factors outside of the online sphere, such as those used within public mental health frameworks. Broad definitions of both radicalisation and public mental health promotion were used, considering the range of very different understandings associated with both terms, and the very varied consequences of such understandings.

To encompass this range of understandings, ‘radicalisation’ was defined by terms that acknowledged the definitional debate around process versus outcome, such as indoctrination, and encompassed far right, far left, conspiratorial, Islamist and other views that have the intent of shifting the status quo. ‘Public mental health promotion’ was defined as incorporating community, school based, and clinical approaches to address mental health concerns and to promote resilience and wellbeing where one of the desired outcomes fell within the parameters of a psychological domain, building on a public mental health promotion orientation for public mental health interventions [16,17].

This review has been conducted as part of the European Union’s Horizon 2020 research and innovation programme¹ focusing on social exclusion and marginalisation as experiential factors in people inclining or turning toward radicalisation, broken down into four public mental health working hypotheses. The first hypothesis concerns spatial formations, stating that *a sense of safety and security, and a sense of belonging, or the lack thereof, can operate as protective or risk factors for extremism*. The second hypothesis concerns identity politics, stating that *identity-building behaviours including ritualised activities and the identification of existential and everyday meaning-making symbols are used for marking in-group and out-group belonging and function to reinforce the identity process by those creating and maintaining online sites. Emotion manipulation techniques are used for triggering hyper arousal responses for users of the site in order to attract and maintain identity at individual and group levels*. The third hypothesis concerns intergenerational change and continuity, stating that *identity and belonging are reinforced by the attempt to reconfigure, weaken, or replace existing nuclear family/clan bonds (where they exist) as well as targeting those lacking such bonds, in order to create new family/clan constellations*. The fourth hypothesis concerns reciprocal radicalisation, stating that *a political shift of governance can bring about new patterns of reciprocal radicalisation (e.g., related to changes in legislation and policies) that can adversely affect health (mental health)*. Set out below is the search strategy, corresponding results and discussion, framed within the four hypotheses outlined here.

2. Materials and Methods

Articles were searched for using Medline (via Ovid), PsycInfo (via Ebscohost) and Web of Science (Core Collection), with the use of Boolean operators and search terms listed in Table 1. This combination of databases has been shown to perform best at achieving efficient and adequate coverage of studies [18]. Additionally, the Cochrane library was searched for clinical interventions. The search resulted in a database of 132 articles, which was deduplicated using the Rayyan systematic review screening software. Articles were then systematically filtered in line with the eligibility and exclusion criteria outlined below. The review followed the guidelines set out by PRISMA [19]. The PRISMA flowchart (Figure 1) summarises the filtering process to exclude articles that did not fulfil the eligibility criteria. Articles were first excluded by title and then abstract, with the remaining excluded through full text. At each point of the exclusion process the articles were assessed against the eligibility criteria. If authors determined that an article failed to meet the eligibility criteria it was excluded.

¹The DRIVE project, *Determining multi-levelled causes and testing intervention designs to reduce radicalisation, extremism, and political violence in north-western Europe through social inclusion*, grant agreement No. 959200. It expresses exclusively the authors' views, not necessarily those of all the DRIVE project Consortium members, and neither the European Commission nor the Research Executive Agency is responsible for any of the information it contains.

Table 1. Search Terms.

Extremism Keywords	Online Keywords	Intervention Keywords	n
Medline via OVID ("Radical Islam*" OR "Islamic Extrem*" OR Radicali* OR "Homegrown Terror*" OR "Homegrown Threat*" OR "Violent Extrem*" OR Jihad* OR Indoctrinat* OR Terrori* OR "White Supremacis*" OR Neo-Nazi OR "Right-wing Extrem*" OR "Left-wing Extrem*" OR "Religious Extrem*" OR Fundamentalis* OR Anti-Semitis* OR Nativis* OR Islamophob* OR Eco-terror* OR "Al Qaida-inspired" OR "ISIS-inspired" OR Anti-Capitalis*).ti,ab. OR terrorism/	("CYBERSPACE" OR "TELECOMMUNICATION systems" OR "INFORMATION technology " OR "INTERNET" OR "VIRTUAL communit*" OR "ELECTRONIC discussion group*" OR "social media" OR "social networking" OR online OR bebo OR facebook OR 3nstagram OR linkedin OR meetup OR pinterest OR reddit OR snapchat OR tumblr OR xing OR twitter OR yelp OR youtube OR TikTok OR gab OR odysee OR telegram OR clubhouse OR BeReal OR Twitter OR WhatsApp OR WeChat OR "Sina Weibo" OR 4Chan).ti,ab. OR internet/ OR social media/ OR online social networking/	("Public mental health" OR "care in the community" OR "mental health service*" OR "educational service*" OR "social service*" OR "public service partnership*" OR "primary care referral" OR "referral pathways" OR "clinical program*" OR "health promotion" OR prevention).ti,ab. OR community mental health services/ OR health promotion/	11
PsycInfo via Ebsco-host TI ("Radical Islam*" OR "Islamic Extrem*" OR Radicali* OR "Homegrown Terror*" OR "Homegrown Threat*" OR "Violent Extrem*" OR Jihad* OR Indoctrinat* OR Terrori* OR "White Supremacis*" OR Neo-Nazi OR "Right-wing Extrem*" OR "Left-wing Extrem*" OR "Religious Extrem*" OR Fundamentalis* OR Anti-Semitis* OR Nativis* OR Islamophob* OR Eco-terror* OR "Al Qaida-inspired" OR "ISIS-inspired" OR Anti-Capitalis*) OR AB ("Radical Islam*" OR "Islamic Extrem*" OR Radicali* OR "Homegrown Terror*" OR "Homegrown Threat*" OR "Violent Extrem*" OR Jihad* OR In doctrinat* OR Terrori* OR "White Supremacis*" OR Neo-Nazi OR "Right-wing Extrem*" OR "Left-wing Extrem*" OR "Religious Extrem*" OR Fundamentalis* OR Anti-Semitis* OR Nativis* OR Islamophob* OR Eco-terror* OR "Al Qaida-inspired" OR "ISIS-inspired" OR Anti-Capitalis*) OR (DE "Terrorism") OR (DE "Extremism")	TI ("CYBERSPACE" OR "TELECOMMUNICATION systems" OR "INFORMATION technology " OR "INTERNET" OR "VIRTUAL communit*" OR "ELECTRONIC discussion group*" OR "social media" OR "social networking" OR online OR bebo OR facebook OR 3nstagram OR linkedin OR meetup OR pinterest OR reddit OR snapchat OR tumblr OR xing OR twitter OR yelp OR youtube OR TikTok OR gab OR odysee OR telegram OR clubhouse OR BeReal OR Twitter OR WhatsApp OR WeChat OR "Sina Weibo" OR 4Chan) OR AB ("CYBERSPACE" OR "TELECOMMUNICATION systems" OR "INFORMATION technology " OR "INTERNET" OR "VIRTUAL communit*" OR "ELECTRONIC discussion group*" OR "social media" OR "social networking" OR online OR bebo OR facebook OR 3nstagram OR linkedin OR meetup OR pinterest OR reddit OR snapchat OR tumblr OR xing OR twitter OR yelp OR youtube OR TikTok OR gab OR odysee OR telegram OR clubhouse OR BeReal OR Twitter OR WhatsApp OR WeChat OR "Sina Weibo" OR 4Chan) OR (DE "Internet") OR (DE "Social	TI ("Public mental health" OR "care in the community" OR "mental health service*" OR "educational service*" OR "social service*" OR "public service partnership*" OR "primary care referral" OR "referral pathways" OR "clinical program*" OR "health promotion" OR prevention) OR AB ("Public mental health" OR "care in the community" OR "mental health service*" OR "educational service*" OR "social service*" OR "public service partnership*" OR "primary care referral" OR "referral pathways" OR "clinical program*" OR "health promotion" OR prevention) OR DE "Public Mental Health" OR DE "Mental Health Services" OR DE "Social Services" OR DE "Health Promotion" AND DE "Prevention" OR DE "Preventive Health Services" OR DE "Preventive Mental Health Services"	27

	Media") OR (DE "Online Social Networks")	
Web of Science (Core Collection)	TS=("Radical Islam*" OR "Islamic Extrem*" OR Radicali* OR "Homegrown Terror*" OR "Homegrown Threat*" OR "Violent Extrem*" OR Jihad* OR Indoctrinat* OR Terrori* OR "White Supremacis*" OR Neo-Nazi OR "Right-wing Extrem*" OR "Left-wing Extrem*" OR "Religious Extrem*" OR Fundamentalis* OR Anti-Semitis* OR Nativis* OR Islamophob* OR Eco-terror* OR "Al Qaida-inspired" OR "ISIS-inspired" OR Anti-Capitalis*)	TS=("Public mental health" OR "care in the community" OR "mental health service*" OR "educational service*" OR "social service*" OR "public service partnership*" OR "primary care referral" OR "referral pathways" OR "clinical program*" OR "health promotion" OR prevention)
Cochrane Library	Radical*, Extrem*, Terrorism, Neo Nazi, terror*, homegrown, jihad, ,indoctrin* supremacis*, right wing, left wing, religious, fundamentalis*anti-semeti*, nativis*, Islam*, Al-Qaida, ISIS, Anti-capitalis*	90
Total		132

2.1. Eligibility Criteria

Articles were limited to peer reviewed studies including reviews, meta-analyses, narrative reviews, pre-post studies, case studies, interventions, cohort studies, cross sectional studies, policy and public service-related work. Articles were included if they contained three elements: firstly, online extremist content (with a main focus on right wing and Islamist extremism, but also a sub focus on other forms of extremist content such as left wing and anti-capitalist content); secondly, mental health and psychological processes; thirdly, an intervention study (see Population, Intervention, Control, Outcome [PICO] criteria below). We defined an intervention study as one in which the researcher “actively interferes with nature – by performing an intervention in some or all study participants – to determine the effect of exposure to the intervention on the natural course of events” [20]. Studies were included if they were written within the last ten years, to account for the rise of social media platforms and extremist content.

2.2. Population, Intervention, Control, Outcome (PICO)

The PICO framework [21] was used to address the public mental health based enquiry arising from this review as outlined below:

Population: Separated by age (young: 11–17-year-old; young adult: 18–25-year-old; adults 25+); separated by milieu (e.g., Islamist, far-right, other).

Intervention: Public mental health, including care in the community, child and adolescent mental health services (CAMHS) and CAMHS-like organisations across different countries, education, social service, public service partnerships, primary care, referral pathways, clinical programmes, health promotion, prevention including school-based interventions.

Control: articles sorted into whether the study used or did not use a control group, to assess the clinical quality of interventions.

Outcomes: Diagnostic outcomes, cognitive, mental health, wellbeing, resilience, psychosocial, social determinants of mental health, learning difficulties, neurodevelopmental conditions, affect and emotional responses.

3. Results

3.1 Data Extraction and Analysis

After duplicates were removed, articles were subjected to a two-stage screening procedure. At the first stage, articles were screened by title and abstract by two authors (RM and EBM). At the second stage, articles were screened through full text by two authors (RM and EBM) and discussed between four authors (RM, EBM,VDM, MN). A data extraction table was then created outlining the PICO variables, details of interventions and each author, journal, year and other publication details. The authors independently retrieved thirty one relevant papers (from the initial 132 papers) according to the inclusion criteria by title and abstract screening (See Figure 1). Following assessment of full text, all retrieved papers were excluded based on the inclusion criteria.

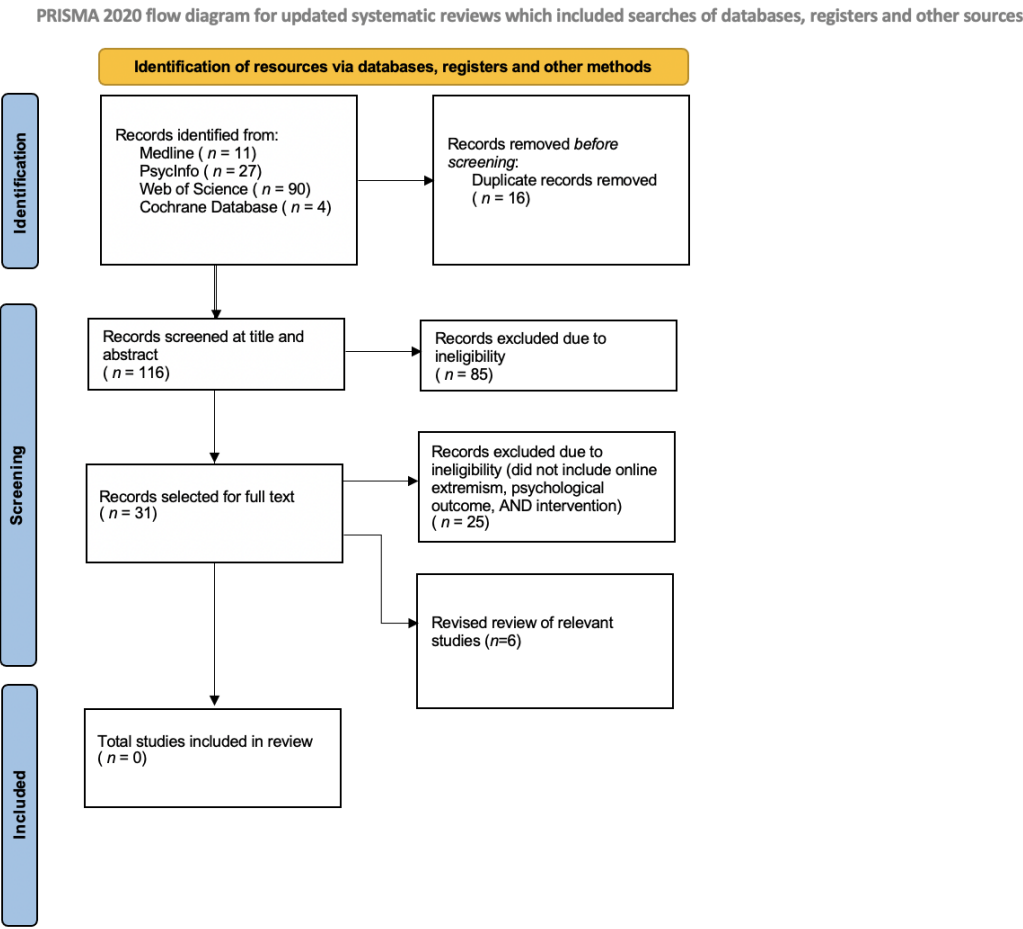


Figure 1. PRISMA Diagram.

No studies were identified that presented public mental health intervention approaches to online extremism. However, the literature search identified six relevant articles, see diagram above, which did not fit the inclusion criteria but identified theoretical relationships between all three elements, online extremism, psychological outcomes, and intervention strategy. Considering the planned review results and the need for attention to the development of interventions to address online radicalisation, we therefore revised the planned review focus to examine these relevant articles to help move forward the development of interventions in this arena of radicalisation. Table 2 provides a summary of the relevant studies before the themed outline.

Table 2. Overview of relevant studies.

Name	Country/ <i>Milieu</i>	Type of article	Summary	Central aim	Central finding/ argument	Reason for relevance/ reason for exclusion
Schmitt et al., 2018	Germany/ <i>Islamist</i> , USA/ <i>Far-right</i>	Information network analysis	An online network analysis of the links between online extremist content and counter extremist messages given that the quantity of extremist messages vastly outnumbers counter messages, both use similar keywords, and automated algorithms may bundle the two types of messages together: counter messages closely or even directly link to extremist content.	Authors used online network analyses to explore what might hinder a successful intervention addressing online radicalisation.	Extremist messages were only two clicks away from counter messaging. The authors suggest that the algorithm filtering and gatekeeping functions directing content and users toward one another, including user amplification through sharing and ‘likes’, as well as the overwhelmingly larger volume of online extremist compared to counter extremist content, together pose almost insurmountable challenges to online interventions targeting extremist content.	Article addressed online extremist content, its relationship with user behaviour and attitudinal shift, and analyses of interventions used. Excluded due to not featuring an intervention but a network analysis of online counter messages. Public mental health approaches might utilise online counter messages as part of an intervention, but no intervention was tested in this article. Rather, several obstacles to counter messaging efficacy were identified.
Rusnalasari et al., 2018	Indonesia/ <i>Islamist</i>	Cross sectional analysis	An analysis of the relationships amongst literacy and belief in and practice of the Indonesian national ideology of Pancasila and literacy in extremist ideological language, with the view of demonstrating that belief and literacy correlate with less vulnerability to online radicalising content: belief and literacy were negatively correlated with vulnerability.	Authors explored if a national ideology warranted testing as an intervention to reduce vulnerability to (or offer protection against) online radicalisation.	National ideology did not seem to reduce vulnerability to or offer protection against online radicalisation.	Article addressed online extremist content, its relationship with language outcomes in the cognitive domain, and theorised the type of intervention that may be useful within education settings. Excluded due to not identifying or testing a specific intervention.
Bouzar & Laurent, 2019	France/ <i>Islamist</i>	Single case study analysis	A qualitative interdisciplinary analysis of the radicalisation of and disengagement intervention with ‘Hamza’, a 15 year old French citizen who attempted several times to leave the country to prepare an attack on France: analysis concludes that Hamza’s life course and related trauma experiences led to radicalisation through the interaction of 3 cumulative processes, emotional, relational and cognitive-ideological.	Authors retrospectively identified the conditions necessary to enable a successful intervention, including the first steps of the intervention.	Argued for the efficacy of a multi-disciplinary intervention that analyses an individual’s life trajectory (rather than only one or two time points) informed by two first steps: i) thematic analyses of semi-structured interviews with parents and the radicalised individual; ii) when permission is granted, and access is legal, thematic analyses of mobile phone and computer records revealing the frequency, content, and patterns of engagement between the individual and the extremist recruiters.	Article addressed online extremist content, its relationship with several psychological domains including affect related trauma and outlined the outcome of an intervention. Excluded as specifics of the intervention were not identified.

Siegel et al., 2019	Global/ <i>Several</i>	Narrative review (book chapter)	A book chapter reviewing pathways to and risk factors for radicalisation, theoretical explanations as to why youth may become radicalised, and recommended intervention approaches and examples in six overlapping arenas (family, school, prison, community, internet, government): review concludes that trauma-informed approaches across the six interacting systems are required.	Authors offered a chapter-length overview on reducing terrorism and preventing radicalisation in six overlapping arenas: family, school, prison, community, internet, and government (the latter referring to diverse services at the international, national, and local levels, depending on country and region, e.g., resource provision to schools, prisoner aftercare, public-private partnerships, financial support services, internet monitoring, law enforcement).	Identified five arenas overlapping with the digital arena in which interventions should be located (family, school, prison, community, government) and argued that two needed approaches are largely absent: trauma informed and resilience promotion.	Article addressed online extremist content, its relationship to trauma, and theoretical areas where interventions may take place. Excluded as example specific interventions were only mentioned and none were tested.
Tremblay, 2020	Global/ <i>Extreme right wing, Far-right</i>	Narrative review (editorial)	An editorial focussing on the alt-right movement, using the terrorist attacks in Christchurch in 2019 as an example: the attack was "A sign of our digital era and social-mediatized gaze", having been live streamed on Facebook and widely shared across the virtual community. The development of inclusive habitats, governance, systems and processes were identified as significant goals for health promotion to foster "peaceful, just and inclusive societies which are free from fear, racism, violation and other violence".	Author provided a very brief high-level analysis focusing on the intersectionality of discrimination and oppression with radicalisation in the digital, political, and social spheres.	Argued for multi-sector partnerships with public mental health promotion approaches to reduce discrimination, oppression, and radicalisation in the digital, political, and social spheres.	Article addressed online extremist content, and areas within public mental health promotion where interventions may take place. Excluded as no specific interventions mentioned or tested.
Schmitt et al., 2021	Germany/ <i>Anti-refugee</i>	Between Subjects design	A study examining the effects of a counterposing intervention with two different narrative structures, one-sided (counter only) or two-sided (extremist and counter) using the persuasion technique of narrative involvement operationalised as two different types of protagonists (approachable or distant/ neutral). The narrative focused on a controversial topic (how to deal with the number of refugees in Germany) and the effect of each narrative structure on attitude change was measured: participants who read the two-sided narrative showed less reactance; the smaller the reactance, the more they felt involved in the	Drawing on findings from the earlier Schmitt et al. (2018) article and theoretical concepts around one sided counter narratives, two sided counter narratives, and narrative involvement, this intervention measured: manipulations, attitude change, freedom threat, and narrative involvement.	Less reactance from a two-sided versus one-sided narrative, that is, from a narrative that included an extremist as well as a counter message. Less reactance was accompanied by increasing narrative involvement (measured as transportation into the narrative and identification with the main character) and self-reported positive attitudinal change toward refugees.	Article addressed online extremist content, its relationship with user behaviour and attitudinal shift, and analyses of the psychological mechanisms involved in mediating the effects of different narrative structures. Excluded as not an intervention but a study that could inform an intervention design using counter messaging.

narrative, which in turn led to more positive attitudes towards refugees; variations in protagonist failed to show an effect.

The literature summarised above is presented below in a three-themed outline: range of online content; who accesses the content; radicalisation and its correlates.

3.2. Range of online content

The six publications summarised above (three articles, a conference paper, a book chapter, and an editorial) outlined a range of online content, mainly shared within social media and video platforms such as Youtube. Analysis of extremist online content varied across the publications, from granular assessment of narrative structure to words associated with terrorism to general descriptions or official definitions as well as content targeting users' existential questions and psychosocial needs. Importantly, all publications within this search noted the interaction or fluidity between the online and offline worlds, with external influences such as education levels or socioeconomic factors most strongly impacting cognitive and behavioural outcomes.

Schmitt et al. [22] conducted an information network analysis to determine the likelihood that users viewing counter extremist videos would also access extremist informational and commentary style videos. They defined online extremist content as that involving a desire to radically, forcefully or violently impose an alternative ideology with totalistic claims stemming from what is considered to be a "true understanding" of the world (p.782). Counter extremist content involved alternative "positive" or "civic education" content with the purpose of steering away from extreme content such as hate speech, conspiracy theory or propaganda (p.783). The analysis included within this study examined two "exemplary" counter-messaging campaigns, #WhatIS, an anti-Islamist platform run by the German Federal Agency of Civic Education, and Life After Hate, an anti-far right platform run by ExitUSA. The authors cited studies indicating that efficacy of counter messaging is mixed, and persuasive mainly with those already expressing doubts (p.783). They went on to describe how online counter messaging's effectiveness is reduced further by the combination of similar keywords being used by both counter and extremist sites, the organising and gatekeeping functions of algorithms that direct content and users toward one another, and the much greater volume of extremist content that exists on platforms (pp.784-786).

Rusnalasari et al. [23] proposed that "vulnerable" internet users, particularly adolescents, access online content pertaining to "individualism, fundamentalism, radicalism, and terrorism" (p.1). Noting that online radicalisation was one of the contributing factors to the Bali bombings of 2002, the authors asserted that increased understanding of online radicalisation processes could address or prevent future instances of violent extremism.

Bouzar & Laurent [24] also focused on the interaction between online content and the needs of internet users. They presented a single subject case study of fifteen-year-old 'Hamza' in France, whose unresolved mourning, sense of societal injustice, and existential questioning were exploited by online ISIS recruiters until he desired martyrdom as entry into a new life. Noting that most ISIS recruits aged 15 – 30 are looking for an idea, a group, and strong emotions, the authors analysed the recruiters' messaging history with Hamza and identified complex emotional and relational strategies that effected cognitive change to achieve group membership, loyalty, and self-sacrifice. The authors presented a calculated, interactive model of online engagement by ISIS recruiters with increasingly extremist online content accessed through and accompanied by frequent communication with recruiters on social media. The conclusion was that a 'perfect storm' occurred involving a vulnerable young person interacting with online recruiters who provided answers to existential questions that family members, school friends and teachers, community and religious leaders, and wider social online and offline systems failed to address.

Siegel et al. [25] reviewed theoretical explanations of radicalisation using a trauma-informed perspective to examine risk factors. They identified overlapping factors in family, school, prison, community, governmental (e.g., resource provision to schools, prisoner aftercare, public-private partnerships, financial support services, internet monitoring, law enforcement) and internet environments that together contribute to radicalisation in different countries (e.g., US, Europe, Australia), rather than simply the content of online material alone. While careful not to attribute causality to trauma, the authors theorised that trauma experiences not met with trauma informed support, combined with family, community, public institutions, governmental policy factors that interacted with online extremist content, and could lead to radicalised beliefs and actions.

Similarly, Tremblay [26] argued that a "vicious interplay" between digital, societal and political spheres contributes to radicalisation, and that public mental health promotion "ought to play a role in addressing the factors contributing to extremism" (p.2). Tremblay in addition cites various public, institutional and social factors that can promote current trends in extreme right-wing ideology.

Schmitt et al. [27] examined whether online content containing characters that the reader can identify and empathise with, i.e., are approachable rather than distant and neutral, was more conducive to cognitive manipulation. The authors focussed on narratives around immigration, refugees and socioeconomic divides and counter messages directed against extremist ideologies, and violent tendencies that exposed the manipulative or propagandic nature of extremist messages. Considering narrative engagement as determinative of how profoundly the content impacts the user, they argued that through narrative involvement a user may temporarily lose connection to reality and escape into the character's world: the more users are transported into a story, the more likely they are to engage in story-consistent beliefs and be susceptible to persuasion, which in turn can deepen or counter radicalised beliefs, depending on the content [27].

3.3. *Who accesses online content*

All six of the publications discussed within this revised review, not satisfying the inclusion criteria of the planned review but summarised due to being most relevant, assumed that online content is widely and increasingly accessed across the general population and acknowledged extensive use by those involved in extremist milieus. Whilst noting that extremist online content can target young users, the articles examined 'who' by considering why and how users engaged with extremist or counter extremist content and the attitudinal and/or behavioural effects of that engagement.

Schmitt et al.'s network analysis of extremist and counter extremist messaging on Youtube noted some behaviours particular to internet users who engage in hate speech, propaganda, violent extremism or conspiracy theories [22]. For example, to spread beliefs, they predominantly rely on social media channels, where messages and ideas can reach a wider audience at a faster rate than in person and filter through networks of social media users. The authors report that messages are often targeted towards younger users through popular media culture such as gaming, music videos and viral videos, often using 'wolf in sheep's clothing' tactics [22]. Rusnalasari et al. examined the effectiveness of a form of civic education in equipping young people to recognise and turn away from extremist online content [23]. This cross-sectional study used data from 193 13–21-year-old participants recruited using purposive sampling through social media links and local contacts who self-reported as active internet users. Extremist ideological literacy levels were correlated with belief in a civic ideology, Pancasila, the Indonesian national ideology of peaceful co-existence among five religions: Islam, Confucianism, Catholicism, Protestantism, and Buddhism. However, the authors found little evidence of protection against radicalisation from belief in Pancasila and literacy in the extremist ideological language used online to promote terrorism. Similarly, the case study by Bouzar & Laurent (2019) plotted the psychosocial journey of a 15-year-old French citizen who accessed not only online extremist content during an existential crisis but interacted online with ISIS recruiters who subsequently convinced him to leave France for Syria. Schmitt et al.'s 2021 analysis of an older group of 405 participants (mean age 40.68, SD = 15.15, recruited via a non-probability access panel or 'convenience pool') found less reactance from a two-sided versus one-sided narrative, that is, from a narrative that included an extremist as well as a counter message [27]. Less reactance was accompanied by increasing narrative involvement (measured as transportation into the narrative and identification with the main character) and self-reported positive attitudinal change toward refugees. Siegel et al. offered a theoretical outline of pathways to radicalisation for young internet users, noting that those who have experienced trauma within various psychosocial events are at higher risk of radicalisation [25]. Tremblay [26] asserted that far-right terrorist events and hate crimes evidence a deeper social and well-being malaise played out in the overlapping digital, societal and political spheres, citing work on the relationship between discrimination and health inequalities by Krieger [28] and the World Health Organization [29]. Racism and discrimination expressed and experienced across all three spheres can create and reinforce ideas of social dominance and oppression. Tremblay [26] noted that associations have been found among social experiences of discrimination or oppression, impaired biological function, and reduced capacities to adapt and cope with social and contextual challenges. Tremblay also cites arguments by Wilkinson and Pickett [30] that as inequity increases, violence and 'perceived threat to pride' increase, which in turn drives radical and extremist narratives and reinforces maladaptive reactions such as oppressive and discriminatory behaviours (p.2). For populations who experience social malaise, Tremblay notes, online content can reinforce radicalised behaviours.

3.4. *Online radicalisation and correlates*

The six publications that were relevant each approached radicalisation from different angles and therefore identified a range of correlates that varied in type and kind: the function and role of social media algorithms; literacy in a national ideology and words encountered online that are associated with terrorism; online recruitment discourses

targeting psychosocial needs; trauma events and experiences in a world with decentralised, ever accessible internet and social media; public mental health promotion research, oppression and the role of the digital sphere; and the narrative structure of online counter messaging. An increasingly hybrid world where lived realities occur simultaneously online and offline involves a wide array of correlates associated with radicalisation, criss-crossing sectors, domains, and conceptual frameworks.

Schmitt et al.'s 2018 article reported on the algorithmic interconnections between counter extremist videos and extremist videos in two 'successful' campaigns that posted counter-messaging videos on YouTube [22]. In doing so, the authors examined the way users can interact with the videos in each campaign. The study showed that extremist content could lead to counter extremist content, if relevant keywords were used within counter extremist content. However, even if the keywords differed, extremist content online could still be accessed within two clicks of a counter message. In contrast, it was unlikely that viewers of counter messaging content would view more *counter* content given the personalisation algorithms that direct similar content toward users, the overwhelmingly larger number of online extremist messages compared with counter messages, and the 'relevance algorithms' that focus on activity level rather than popularity metrics (p.798).

To examine the interactions among literacy and belief in a national civic ideology (Pancasila), literacy in extremist ideological language, and vulnerability to extremist recruiters, Rusnalasari et al. [23] used a cross-sectional design with an online survey to collect responses to: i) nineteen questions measuring literacy about extreme words that could be used to promote terrorist ideologies (3-5 items per each concept - individualism, conservatism, fundamentalism, radicalism, terrorism), ii) five questions measuring participants' understanding of the civic ideology (Pancasila), and iii) 'a few questions' collecting demographic information (p.3). Prior to correlation analyses, the questionnaires were validated, and reliability tested using logical regression with a finding of 95% confidence levels (p.3). The authors found higher literacy levels about extreme ideologies were interconnected and correlated with higher understandings of Pancasila. Despite this finding, 'several' further questions about how to put into practice the belief of peaceful co-existence elicited 70% 'wrong' answers, which the authors hypothesised indicated being 'vulnerable' to 'decide the wrong reaction' that would tend to 'change into action of terrorism' (p.5).

Bouzar & Laurent [24] conducted qualitative thematic analyses of semi-directive interviews with the subject, 'Hamza', and his parents. Given permission by the parents and Hamza, and with legal approval, the interdisciplinary disengagement team conducted an analysis of his online engagement with ISIS on his computer and mobile phone, representing unusual access to the videos viewed and shared within the IS group. These first steps of the intervention (interviews and analysis of online engagement) produced combined analyses that revealed the emotional, relational, psychological, social stages, and reasoning processes underlying his radicalisation that began online. The subsequent steps of the intervention continued with this multidisciplinary approach. However, the article does not describe in detail the subsequent steps of the intervention that were used with Hamza but indicates that there was continued engagement for a period of time. The article focuses on the benefits of and need for an interdisciplinary approach using thematic analysis, and, where ethically viable, access to mobile phone and computer records to analyse engagement with recruiters, as part of an intervention that will "untie" violent extremist beliefs ([24], p.664). As researchers and practitioners, the authors emphasise that to be authoritative the extremist recruitment discourse makes a difference in the young person's life and therefore the disengagement discourses *also* must make a difference in the young person's life. Methodologically, the disengagement intervention steps are depicted as targeting the explicit discourse of the online recruitment process and the implicit motivations of the young person that were identified during the thematic analyses.

Siegel et al. [25] did not report on a specific intervention but recommended intervention parameters. Reviewing theoretical explanations of adolescent and young adult radicalisation, they used a trauma-informed perspective to examine risk factors for radicalisation. They identified family, schools, prison, community, internet, and government programmes and services (for example counter terrorism strategies) as six inter-relating arenas of radicalisation and therefore entry-points for intervention. Radicalisation was defined as the process of adopting an extremist belief system, including 'the willingness to use, support or facilitate violence, as a method to effect societal change' (p.392, quoting Allen [31], p.4). Their focus was particularly on radicalisation as a precursor to terrorist activities, understood as 'any action intended to cause death or serious bodily harm to civilians or non-combatants with the purpose of intimidating a population or compelling a government or international organisation to do or abstain from an act' (p.392, quoting the United Nations [32]). While acknowledging that no clear profile or universal list of risk factors for radicalisation or terrorism exists, Siegel et al. asserted that examination of the radicalisation process often reveals trauma events and experiences, and a blurry border between widely held ideologies and radicalised belief systems that can lead to violent acts. They observed that most terrorist organisations employ youth aged 15-22 years old in some capacity, recruiting by

offering a clear identity, belonging and adventure. Offering an overview of the many theories about why some move from cognitive to violent radicalisation, they commented that none directly touch on the role of trauma in the radicalisation process and called for research on the roles of trauma in radicalisation and of resilience promotion in preventing radicalisation, including community based public (mental) health promotion. As decentralised systems, impossible to control, censor or restrict, and with ever growing accessibility, the internet and social media was named one of the most significant resources in the radicalisation processes. Counter-messaging, incentivising reporting, increasing digital literacy, and online dialogue with extremists by trained volunteer scholars were cited as example interventions ([25] p.408-9).

Tremblay [26] argued that public health (including mental health) promotion can support coordinated multi-sectoral models to empower those who are oppressed and to meet oppressors “where they are” ([26], p.3). These efforts should include aims to develop skills in critical thinking, digital literacy, and promote inclusive social norms. Tremblay concludes that public health, including mental health, promotion must prioritise research (including interventions) on the associations among inclusion, fairness, and health within the economic, political, social and cultural landscapes. Calling for public health promotion involvement in this field, Tremblay noted that immigration is instrumentalised by extremist ideologies and nationalist agendas promulgated in digital, social and political spheres, represented, for example, by the “Great replacement theory” ([26], p.1,2).

Schmitt et al. [22] compared the effectiveness of two types of online ‘counter messaging’ narrative structures in changing user attitudes (and, theoretically, behaviour). The volume of extremist messaging vastly outnumbers counter messaging, and the latter sometimes contains extremist content in order to deconstruct it. But this can make the user vulnerable to extremist content through the algorithm gatekeeping, filtering, and amplification mechanisms, which can lead users to extremist content, without the user realising it. Given this potential risk, the authors sought to determine if one or two-sided narratives were more effective at changing attitudes, the former presenting only a counter message, the latter presenting both the counter message and the message being countered. The study was presented as the first of its kind, synthesising findings from several previous studies (e.g., [33-37]) and from psychological research used in advertising (e.g., [38-39]). The authors also examined the role of reactance, defined as “a physiological arousal in reaction to a certain external stimulus which occurs if people feel that their freedom of opinion is being threatened” ([27] p.58). Defining freedom threat as an essential condition of reactance, and an antecedent to further affective and cognitive aspects of experiencing reactance ([27] p.58), the study argued that elevated reactance renders a reader less likely to accept a persuasive message. In contrast, less reactance supports narrative involvement and lowers the risk that a reader will feel that they are being forced toward a particular position. In this study, four alternative texts were presented to participants. Each narrative presented a description of a young woman named Lena, who has strong positive attitudes towards refugees in Germany. Lena meets her long-term friend Anne. By chance, they start talking about the refugee crisis in Germany. As proposed by Cohen and colleagues [34], in the ease of identification condition, one character, Lena, is portrayed more positive, virtuous, and described in detail, whereas these attributes are missing from the other character (without portraying her negatively). Two political opinions about refugees were presented: one character (Lena) expressed pro-refugee arguments whereas the other character (Anne) presented contra-refugee attitudes. In this condition, the two friends start to debate about the topic. Arguments alternated between pro and contra asylum seekers. The debate becomes increasingly emotional and ends with the suggestion to talk about something else to prevent a serious fight. In contrast, the one-sided narrative presented only the pro-asylum seekers arguments of Lena, whereas Anne is a neutral audience to Lena’s arguments resulting in no emotional debate.

4. Discussion

The planned review found no individual, family based, community based or institutionally based public mental health intervention studies related to online extremism. We set out to investigate the nature of online extremist content, the demographics of individuals who access extremist content, and interventions focusing on psychological domains using a public mental health approach. We searched for peer reviewed studies, with ‘study’ defined broadly to include reviews, meta-analyses, narrative reviews, pre-post studies, case studies, interventions, cohort studies, cross-sectional studies, policy and public service-related work. Inclusion depended on the content containing three criteria: (i) online extremist content with a main focus on right wing and Islamist extremism, but possibly a sub focus on other forms of extremist content, e.g., anti-capitalist extremist content; (ii) mental health and psychological processes; and (iii) description of an intervention using the PICO criteria (Population, Intervention, Control, Outcome). Intervention was defined as one in which the researcher “actively interferes with nature – by performing an intervention in some or all study

participants – to determine the effect of exposure to the intervention on the natural course of events” ([20], p.137). Searches on multiple data bases found 132 articles that were double screened down to 31 articles with no publication fitting all 3 criteria. A search of the Cochrane Library for high-quality controlled trials, randomised or quasi- randomised, did not yield any registered studies. The revised review included six publications of some relevance that did not fully fit the planned eligibility criteria. These publications were presented here to provide a sample of relevant literature addressing online extremism through varied methodologies (e.g., network analysis, cross-sectional study) or approaches (e.g., international overview, editorial argument) and linked in some way with public mental health approaches.

The studies that were relevant, although they did not fulfil completely the original criteria, identified challenges to addressing online radicalisation and extremism. Two of the studies focused on the hurdles related to designing and disseminating online counter extremist messages that are persuasive and do not amplify extremist content or inadvertently lead users to extremist content [22,27]. A third showed that high literacy levels about extremist language and a nationalist ideology of peaceful coexistence do not reduce vulnerability to online extremist views [23]. A single subject case study noted that most disengagement programmes view radicalisation processes from a single time-point and one disciplinary perspective, but the emotional, relational, psychological, and social tactics that recruiters use to target potential recruits with online messaging and video content require a multi-disciplinary life trajectory approach to enable successful disengagement [24]. The book chapter review of theoretical explanations of radicalisation recruitment and interventions described the internet as decentralised, ever more accessible, impossible to control, censor or restrict and online spaces as where help is often sought for trauma events and experiences that are embedded in many descriptions of radicalisation [25] - and indeed that is what Bouzar and Laurent [24] illustrate in their single case study. The editorial noted the interplay among digital, societal, and political spheres in radicalisation and extremism, arguing for public (mental health) promotion involvement in addressing the social, economic, and psychosocial factors that contribute to experiences of oppression and oppressive behaviours toward others [26]. All highlighted the interaction among psychological factors, life experiences, and the wider social, economic, health, cultural and political context. In addition to being culturally informed and contextually appropriate, the findings from this review emphasise that effective evidence-based wellbeing and health promotion, prevention, and intervention programmes will be part of multi-sector, multi-disciplinary, multi-agency approaches and also provide more granular guidance for addressing online extremism.

Spatial formations, as stated in the first public mental health working hypothesis presented in the Introduction, contain spaces where, “*a sense of safety and security, and a sense of belonging, or the lack thereof, can operate as protective or risk factors for extremism*”. Such spatial formations are found within the online world through online groups and communities that may contribute to group-think or ideological polarisation. Analysing scraped data from a white supremacist online forum, Stormfront.org, Gregory & Piff [40] found that both cognitive complexity and style matching decreased as engagement increased, indicating increased ideological polarisation rather than the deindividuation of groupthink. Studying online video games, Robinson & Whittaker [41] argue that interactive gameplay and the use of iconography such as Nazi memorabilia creates conditions of belonging, in which young players can turn to extremist ideology. While *spatial formation* is not mentioned specifically in any of the articles, the single subject case study outlines in great detail the formation of identities within the spatial intersection of online and offline engagement [24]. Similarly, the editorial By Tremblay [26] notes the interplay among digital, societal, and political spheres in which social inequalities, social determinants of health, and economic, political, social, and cultural landscapes create contexts of oppression, discrimination and deeper social and wellbeing malaise conducive to extremist ideologies and nationalist agendas that circulate. Additionally, different types of media and online content are used by individuals to access extremist material. The literature both within this review and otherwise (see, e.g. [14]) identify social media, peer to peer, video hosting and collaborative platforms as conduits to extremist material, but there appears to be a growing number of radicalised individuals amongst younger users of short-attention video platforms such as Tik Tok, gaming platforms such as Twitch, or online videogames [40]. Tik Tok for example is notable for its lack of enforcement around community guidelines and has been implicated in the rise of alt-right sentiment amongst young people [14] alongside Islamist content accessed by different young populations [41].

Prolonged exposure to social media may impact long term mental health outcomes, including stress related, disordered eating related, and cognitive and attention deficit related outcomes [42], and mental health services are increasingly incorporating the long-term addictive and affect-related effects of social media into treatment plans for adolescents [43,44]. With this in mind the question can be raised as to the purpose of interventions within this field. For the two Schmitt et al. articles ([22,27] each with the same lead author but different co-authors) that appeared within this

search, the purpose of intervening through counter messaging was to effect behaviour change and persuade individuals to believe an alternative viewpoint. Within these articles, a two-sided counter extremist narrative appeared to be more effective at persuasion rather than one-sided narratives [27] alongside facilitating identification with a main character, which transports the user into a different narrative. However, as the authors note, even carefully designed two-sided narratives may fail to divert users from radicalising and extremist content, or reduce the impact of such content, due to a number of reasons. Firstly: a greater quantity of online extremist content exists which drowns out the 'counter-voice'; secondly, counter messages need to use the same wording, catchphrases or conspiracy content to attract viewers; thirdly, approaches such as humour or satire may be misunderstood. Counter messaging may be unproductive as it may exacerbate already held views through reinforcing the message, but it can also be unethical if used by policy makers to fulfil a particular narrative. There are, as Hurlow et al. [6] point out, ethical concerns regarding "the requirement that we monitor and report all unacceptable thoughts" (p.162). The story telling technique employed by Schmitt and colleagues could be employed within video content services. One area that could be exploited is within 'nanolearning', in which retention times are seen to be superior to long form video content [45]. While counter messages like those used by Schmitt and colleagues could be used in public mental health promotion interventions, it is not clear, for instance, who should define what constitutes a 'radical' message, whether public mental health bodies should promote political attitudinal shifts, or whether those already on video sharing platforms should be employed in the dissemination of counter messaging content.

Identity politics, as in the second public mental health working hypothesis, involve "identity-building behaviours including ritualised activities and the identification of existential and everyday meaning-making symbols are used for marking in-group and out-group belonging and function to reinforce the identity process by those creating and maintaining online sites. Emotion manipulation techniques are used for triggering hyper arousal responses for users of the site in order to attract and maintain identity at individual and group levels". The identity politics of exclusion, discrimination, and marginalisation are mentioned explicitly within the outlined literature, with particular focus on the interaction between the online and offline worlds, particularly external influences such as education, community, and socio-economic factors [25,26]. The editorial by Tremblay [26] outlines the impact of these external factors on far-right violence, with mention of the Christchurch Mosque shootings in 2019. In the single subject case study by Bouzar & Laurent [24], Hamza's experiences of marginalisation and discrimination within the family (for example not being allowed to mourn the death of his grandfather) compounded with experiences of marginalisation and discrimination at the community and system level (for example being bullied and othered at school for being ethnically different) thereby created a fragility within his sense of identity – which ultimately was exploited by ISIS recruiters. As argued by Bouzar and Laurent [24], interventions effective at disrupting the radicalisation process need to be interdisciplinary and multi-sector to engage at the individual, family, community, and system levels.

Intergenerational change and continuity, as in the third public mental health working hypothesis, recognised that "identity and belonging is reinforced by the attempt to reconfigure, weaken or replace existing nuclear family/clan bonds (where they exist) as well as targeting those lacking such bonds, in order to create new family/clan constellations". Intergenerational factors played explicit and important roles in the single subject case study [24]. Even during the disengagement process, the authors report that the parents and son never discussed how difficult it was for him to be disallowed from visiting his dying grandfather in hospital, despite their closeness, and from attending the funeral in Algeria. The parents thought they were protecting him from the shock of seeing a loved one covered in tubes and machines, but the son needed to participate in the farewell and mourning rituals to experience closure. The unresolved anger and despair felt by the son again created identity and belonging fragilities, as well as existential questions about life and death. Not finding help in the local mosque where these topics were not discussed, he turned to the internet for answers where online IS recruiters provided answers based on a distorted version of Islam. Moreover, he came to view his family members as unfaithful, since they failed to teach him the true Muslim faith and aspired to ensure his family's place in paradise through his own martyrdom.

Reciprocal radicalisation as stated in the fourth public mental health working hypothesis, represents "a political shift of governance [that] can bring about new patterns of reciprocal radicalisation (e.g., related to changes in legislation and policies) that can adversely affect health (mental health)". Empirical findings related to the 2015 shifting of immigration and asylum policies, for example, in the Swedish case studied in a recent Horizon 2020 project, provide an example of multi-level societal consequences brought about by restrictive policies not only for refugees, asylum seekers and their families but in the larger society as well. In a short span of time, changes in labels and societal perceptions of specific immigrant groups by the majority culture and within immigrant subcultures, and from immigrant groups to the majority culture, transformed from open and supportive to negative and adversarial. These changes were evident

across fluid, hybrid online offline spaces, creating and exacerbating intergroup tensions [46]. This is also illustrated within the Bouzar & Laurent single subject case study [24], in which both domestic and global politics play a role in the pathway to engaging in radicalised discourse. They note that Hamza is willing to justify violent and vigilante behaviour towards French citizens because of violent and graphic content shown to him by an ISIS recruiter. Attitudinal shift is seen when Hamza moves from wanting to help the child orphaned by the air strike, to being ready to commit to killing French citizens on the child's behalf. Similarly, Tremblay [26] notes that oppressive experiences can create oppressive behaviours.

The existence and efficacy of a public mental health intervention approach to online political radicalisation is a topic for future research. The present systematic review searched for a wide range of public mental health intervention approaches in community, education, health and other settings, but none were found that matched the inclusion criteria. In the revised review of six articles examined here, one article reported on an attitudinal change study [27] that examined the psychological mechanisms involved in counter and two-sided narrative designs that could be embedded within extremist content on video hosting sites such as Youtube. Whilst the two-sided design was found to be more effective within the metrics used, the authors noted the inherent limits of online counter-messaging given the much larger volume of extremist content and algorithmic filtering and gate-keeping functions. This study could be used to inform the design of public mental health promotion interventions incorporating counter-messaging, but this raises questions about what public mental health promotion involves, and whether a whole system approach (for example through increased access to education, employment, housing, life experiences, healthy relationships within family and community settings, and greater structural and political security) could address radicalisation. The current review highlights a gap in this field. There is a need for further research into integrated ecosocial approaches to resilience promotion and radicalisation prevention that focus on nested interconnections among individual, family, community and structural levels and include the fluidity of online and offline experiences.

4.1. Potential biases and errors in the review process

This systematic review is the first to consider public mental health approaches to online radicalisation. This is an emerging public mental health field. A comprehensive three-step systematic search in three databases was performed. The systematic search strategy was undertaken by a specialist librarian (VP). This minimised the risk of missing potentially relevant studies during the search process. Paper retrieval was done in two phases independently by two reviewers. Any disagreement was resolved by discussion between the reviewers thereby minimising the risk of missing potentially relevant primary study during paper retrieval.

5. Conclusions

There is a paucity of data within this field. Whilst six articles were identified that had some relevance to this area, it appears not to be known through peer reviewed evidence what public mental health approaches have been used to address online radicalisation. Although this is an emerging public mental health field, the results reported here demonstrate the need for further research in this area. Further research may utilise a realist or rapid evidence approach in which grey literature and items in the public domain may be included within the review protocol.

Author Contributions: Conceptualization, RM, VDM, and EBM; methodology, RM, VDM, MN, VP and EBM; formal analysis, RM, VDM, and EBM.; investigation, RM, VDM, and EBM; writing—original draft preparation RM, VDM, MN, HL, and EBM.; writing—review and editing, RM, VDM, MN, HL, VP and EBM; supervision, VDM, EBM; project administration, RM, VDM, EBM; funding acquisition, VDM, EBM. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Union's Horizon 2020 research and innovation programme: The DRIVE project, Determining multi-levelled causes and testing intervention designs to reduce radicalisation, extremism, and political violence in north-western Europe through social inclusion, grant agreement No. 959200.

Acknowledgments: All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and NIHR Applied Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gill, P., Clemmow, C., Hetzel, F., Rottweiler, B., Salman, N., Van Der Vegt, I. & Corner, E. Systematic review of mental health problems and violent extremism. *The Journal of Forensic Psychiatry & Psychology*, **2021** 32(1), 51-78.
2. Pathé, M. T., Haworth, D. J., Goodwin, T. A., Holman, A. G., Amos, S. J., Winterbourne, P., & Day, L. Establishing a joint agency response to the threat of lone-actor grievance-fuelled violence. *The Journal of Forensic Psychiatry & Psychology*, **2018** 29(1), 37-52.
3. Augestad Knudsen, R. Measuring radicalisation: Risk assessment conceptualisations and practice in England and Wales. *Behavioral Sciences of Terrorism and Political Aggression*, **2020** 12(1), 37-54.
4. Bhui, K.. Flash, the emperor and policies without evidence: counter-terrorism measures destined for failure and societally divisive. *BJPsych bulletin*, **2016** 40(2), 82-84.
5. Summerfield, D. Mandating doctors to attend counter-terrorism workshops is medically unethical. *BJPsych Bulletin*, **2016** 40(2), 87-88.
6. Hurlow, J., Wilson, S., & James, D. V. Protesting loudly about Prevent is popular but is it informed and sensible? *BJPsych Bulletin*, **2016** 40(3), 162-163
7. Koehler, D. Family counselling, de-radicalization and counter-terrorism: The Danish and German programs in context. In *Countering violent extremism: Developing an evidence-base for policy and practice* **2015** (2015)129-138.
8. NIHR Conceptual framework for Public Mental Health. Accessed on 13.01.2023 at: <https://www.publicmentalhealth.co.uk> **2023**
9. NHS Guidance for mental health services in exercising duties to safeguard people from the risk of radicalisation. Accessed on 13.01.2023 at: <https://www.england.nhs.uk/wp-content/uploads/2017/11/prevent-mental-health-guidance.pdf> **2017**
10. Bello, W. F. Counterrevolution: The global rise of the far right. *Fernwood Publishing*. **2019**
11. Moskalenko, S., González, J. F. G., Kates, N., & Morton, J. Incel ideology, radicalization and mental health: A survey study. *The Journal of Intelligence, Conflict, and Warfare*, **2022** 4(3), 1-29.
12. Marwick, A., Clancy, B., & Furl, K.. Far-Right Online Radicalization: A Review of the Literature. *The Bulletin of Technology & Public Life*. **2022**
13. Nilan, P. Young people and the far right. Springer Nature. **2021**
14. Boucher, V.. Down the TikTok Rabbit Hole: Testing the TikTok Algorithm's Contribution to Right Wing Extremist Radicalization *Doctoral dissertation*. **2022**
15. Isaak, J., & Hanna, M. J.. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, **2018** 51(8), 56-59.
16. Fledderus, M., Bohlmeijer, E. T., Smit, F., & Westerhof, G. J. Mental health promotion as a new goal in public mental health care: A randomized controlled trial of an intervention enhancing psychological flexibility. *American journal of public health*, **2010** 100(12), 2372-2372.
17. Boyd-McMillan, E., & DeMarinis, V. Learning Passport: Curriculum Framework (IC-ADAPT SEL high level programme design). **2020**
18. Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic reviews*, **2017** 6(1), 1-12
19. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. Research methodes and reporting. *Bmj*, **2009** 8, 332-336.
20. Aggarwal, R., & Ranganathan, P. Study designs: Part 4—interventional studies. *Perspectives in clinical research*, **2019** 10(3), 137.
21. Huang, X., Lin, J., & Demner-Fushman, D. Evaluation of PICO as a knowledge representation for clinical questions. In *AMIA annual symposium proceedings* **2006** (2006, p. 359). American Medical Informatics Association.

22. Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. Counter-messages as prevention or promotion of extremism?! The potential role of YouTube: recommendation algorithms. *Journal of communication*, **2018** 68(4), 780-808.
23. Rusnalasari, Z. D., Algristian, H., Alfath, T. P., Arumsari, A. D., & Inayati, I. Students vulnerability and literacy analysis terrorism ideology prevention. In *Journal of Physics: Conference Series* **2018** (Vol. 1028, No. 1, p. 012089). IOP Publishing.
24. Bouzar, D., & Laurent, G. The importance of interdisciplinarity to deal with the complexity of the radicalization of a young person. In *Annales Medico-Psychologiques* **2019** (Vol. 177, No. 7, pp. 663-674).
25. Siegel, A., Brickman, S., Goldberg, Z., & Pat-Horenczyk, R. Preventing future terrorism: Intervening on youth radicalization. *An International Perspective on Disasters and Children's Mental Health*, **2019** 391-418.
26. Tremblay, M. C. The wicked interplay of hate rhetoric, politics and the internet: what can health promotion do to counter right-wing extremism?. *Health Promotion International*, **2020** 35(1), 1-4.
27. Schmitt, J. B., Caspari, C., Wulf, T., Bloch, C., & Rieger, D. Two sides of the same coin? The persuasiveness of one-sided vs. two-sided narratives in the context of radicalization prevention. *SCM Studies in Communication and Media*, **2021** 10(1), 48-71
28. Krieger, N. Discrimination and health inequities. *International Journal of Health Services: Planning, Administration, Evaluation*, **2014** 44, 643-710.
29. WHO Commission on Social Determinants of Health, & World Health Organization. Closing the gap in a generation: health equity through action on the social determinants of health: Commission on Social Determinants of Health final report. *World Health Organization*. **2008**
30. Wilkinson, R. and Pickett, K. The spirit level. Why equality is better for everyone. London: *Penguin*. **2010**
31. Allen C.E. Threat of Islamic radicalization to the homeland. Testimony before the US Senate Committee on Homeland Security and Government Affairs. *Washington, DC: United States Senate*. **2007**
32. United Nations General Assembly. *59th Session, A/59/565*. <http://hrlibrary.umn.edu/instreet/report.pdf>. **2004**
33. Braddock, K., & Dillard, J. P. Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behavior. *Communication Monographs*, **2016** 83(4), 446-467
34. Cohen, J. Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society*, **2001** 4(3), 245-264.
35. Moyer-Gusé, Jain, P., & Chung, A. H. Reinforcement or reactance? Examining the effect of an explicit persuasive appeal following an entertainment-education narrative. *Journal of Communication*, **2013** 62(6), 1010-1027.
36. Cohen, J. Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society*, **2001** 4(3), 245-264
37. Green, M. C., & Brock, T. C.. The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, **2000** 79(5), 701-721
38. Allen, M. Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication*, **1991** 55(4), 390-404.
39. Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. Experiments on mass communication (*Studies in social psychology in World War II*, Vol. 3.). **1948** Princeton: Princeton University Press.
40. Gregory AL, Piff, P.K. Finding uncommon ground: Extremist online forum engagement predicts integrative complexity. *PLoS ONE* **2021** 16(1): e0245651
41. Robinson, N., & Whittaker, J. Playing for Hate? Extremism, terrorism, and videogames. *Studies in Conflict & Terrorism*, **2020** 1-36.
42. Zhao, N., & Zhou, G. Social media use and mental health during the COVID-19 pandemic: Moderator role of disaster stressor and mediator role of negative affect. *Applied Psychology: Health and Well-Being*, **2020** 12(4), 1019-1038.
43. Leong, L. Y., Hew, T. S., Ooi, K. B., Lee, V. H., & Hew, J. J. A hybrid SEM-neural network analysis of social media addiction. *Expert Systems with Applications*, **2019** 133, 296-316.

44. Tutgun-Ünal, A., & Deniz, L. Development of the social media addiction scale. *AJIT-e: Bilişim Teknolojileri Online Dergisi*, **2015** 6(21), 51-70.
45. Garcia, M. B., Juanatas, I. C., & Juanatas, R. A. . TikTok as a Knowledge Source for Programming Learners: a New Form of Nanolearning?. In *2022 10th International Conference on Information and Education Technology (ICIET) 2022* (pp. 219-223). IEEE.
46. Cetrez, Ö., DeMarinis, V., Pettersson, J., & Shakra, M.. (2020). *Integration: Policies, Practices, and Experiences, Sweden Country Report. Working papers Global Migration: Consequences and Responses*, **2020** EU Horizon 2020 Country Report for RESPOND project. Uppsala University.