

Conspiracy, Misinformation, Radicalisation: Understanding the online pathway to indoctrination and opportunities for intervention

By Emily Booth^a, Jooyoung Lee^b, Marian-Andrei RizoIU^b, and Hany Farid^c

^a*Faculty of Arts and Social Sciences, University of Technology Sydney, Sydney, Australia*

^b*The Data Science Institute, University of Technology Sydney, Sydney, Australia*

^c*School of Information, University of California, Berkeley, United States*

Emily Booth (Corresponding Author)

Dr Emily Booth is a researcher at the University of Technology Sydney, exploring young adults' reading practices, their encounters with online misinformation, Australian high school curriculum design, and diversity in Australian book publishing. She is a founding member of the Executive Board of the international YA Studies Association (YASA).

Email: Emily.Booth@uts.edu.au | ORCID: 0000-0002-8154-6747

Mail address: School of Communication, University of Technology Sydney, Ultimo, NSW 2007, Australia.

Jooyoung Lee

Dr Jooyoung Lee is a Research Fellow at University of Technology Sydney. Her research includes investigating patterns of online misinformation and designing strategy possibilities to proliferate factual information. Her previous research focused on trust and reputation on online social networks and related issues including link prediction and community search problems.

Email: Jooyoung.Lee@uts.edu.au

Mail address: UTS Data Science Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia.

Marian-Andrei RizoIU

Dr Marian-Andrei RizoIU is a Associate Professor leading the Behavioral Data Science lab at the University of Technology Sydney. His interdisciplinary research crosses computer and social sciences, blending psycholinguistics, digital communication and stochastic modelling to understand human attention dynamics in the online environment, the emergence of influence and opinion polarization.

Email: Marian-Andrei.RizoIU@uts.edu.au | ORCID: 0000-0003-0381-669X

Mail address: UTS Data Science Institute, University of Technology Sydney, Ultimo, NSW 2007, Australia.

Hany Farid

Professor Hany Farid is a professor at the University of California, Berkeley. He is a member of institutes including the Berkeley Artificial Intelligence Lab, and is a senior faculty advisor for the Center for Long-Term Cybersecurity. His research focusses on digital forensics, forensic science, misinformation, image analysis, and human perception.

Email: hfarid@berkeley.edu

Mail address: School of Information, University of California, Berkeley, CA 94720, USA.

Conspiracy, Misinformation, Radicalisation:

Understanding the online pathway to indoctrination and opportunities for intervention

Abstract

In response to the rise of various fringe movements in recent years, from antivaxxers to QAnon, there has been increased public and scholarly attention to misinformation and conspiracy theories and the online communities that produce them. However, efforts at understanding the radicalisation process largely focusses on those who go on to commit violent crimes. This paper draws on three waves of research exploring the experiences of current and former individuals involved in fringe communities, including the different stages of investment they progressed through, and ultimately, what made people leave. We propose a pathway model for understanding contemporary online radicalisation, including potential interventions that could be safely made at each stage. Insight into the experience of being immersed in these communities is essential for engaging with these people empathetically, and therefore preventing both the emergence of violent terrorists and protecting vulnerable people from being drawn into these communities.

Keywords: radicalisation, misinformation, conspiracy theories, vaccines, QAnon

Manuscript word length (excluding abstract): 8004

Introduction

Research has identified that those most susceptible to conspiracy theories, misinformation, and other extremist views are those who are experiencing high levels of uncertainty and insecurity; sometimes in addition to a sense of distrust in, or alienation from, authorities and the world around them (Abalakina-Paap et al., 1999; Grzesiak-Feldman, 2013; Klausen, et al., 2016; Winter and Feixas, 2019). Yet, most people who are drawn into these views do not commit dramatic acts of violence and attitudes towards violence vary even within the same extremist community (Kruglanski, et al., 2022: 1; Vergani, Barton, Iqbal, 2017). Given the devastation caused by incidents like the Wiemballa shooting in Australia in December 2022 in which conspiracy actors murdered several young police officers (Gillespie and McGowan, 2022), it is understandable that governments and researchers alike have prioritised investigating these more extreme cohorts. However, non-violent radicalised individuals still experience negative impacts due to their involvement in fringe communities. For the majority of people who adopt these views, it is their own lives and families that suffer; for example, anti-vaccination parents whose decisions lead to the death of their child from a preventable illness (Evelyn, 2020). On a larger scale, anti-vaccination trends can lead to the loss of herd immunity against easily preventable diseases, such as Measles, which pose significant risk to the wider population who are made newly vulnerable (Ryan, 2021; Attwell et al., 2021). The slow creep of these illnesses back into our societies has the potential to lead to considerable loss of life; and while this is less immediately shocking than the violence of a mass-shooter, it is no less threatening. Furthermore, even non-violent radicalised individuals who do not adopt the most extreme of views can still play a role in disseminating and engaging with misinformation, which sustains the online cultures that will eventually produce violent terrorists.

Understanding how individuals become radicalised and what interventions can be safely made is essential for preventing both the emergence of violent terrorists and protecting vulnerable people from being drawn into these communities. Individuals with major life uncertainties are more prone to being drawn into by conspiracy theories and misinformation (Kruglanski et al., 2022), and the extent to which people are likely to believe in misinformation is greatly influenced by their existing beliefs (Dahlke and Hancock, 2023: 6). With this view, susceptibility to misinformation can be understood as a personal issue first, and an information issue second. As Gerrand notes in relation to violent extremism, the primary reason for joining such groups is a desire to belong (2022: 2). Yet, as witnessed during the ongoing COVID-19 pandemic, conspiracy theories and misinformation can fuel distrust in governments and other trusted authority figures to which people would ordinarily turn to for reassurance, such as doctors. Distrust in these figures normally responsible for the information and support that alleviates anxieties about life can lead to alienation from essential authoritative information (McCauley and Moskalenko, 2008). Likewise, friends and family members who respond with disapproval or mockery to the expression of fringe views can inadvertently push a person to become further radicalised (Scrivens et al., 2022: 669). Distrust as a barrier to potentially positive intervention is therefore not limited to governments and other authority figures.

More indirect efforts to reach people have also been trialled. For people who are exposed to misinformation, intervention through fact-checking can, to some extent, help to better inform people so they can form views based on reliable sources (Walter et al., 2020). Since the COVID-19 pandemic, there has been some increase in social media platforms fact-checking posts, although research has demonstrated that the implementation of this intervention is limited (Xue et al., 2022). However, fact-checking and other approaches that aim to educate the general public can only be effective if the individual trusts and

understands the fact-checking authority, and such an approach does not necessarily address the personal issue at the root of a person's fringe beliefs. Likewise, Thorburn has noted in his study of incel and far-right communities that some influencer content by mental health experts is mentioned by recovering men as aiding in their recovery (2023a: 11). However, when distrust in "experts" is a core component of a fringe community's views, as is the case for many online conspiracy and misinformation groups, such authority may undermine its potential impact. Therefore, there is a need for a more diverse interventions toolkit.

This paper presents a pathway in which we aim to depict a six-stage process by which people become increasingly radicalised online as well as opportunities for interventions at each stage. Many existing models or pathways in our literature search were highly tailored to Jihadist radicalisation or "lone wolf" terrorists springing from incel communities (see Green et al., 2023). While insightful, the actions described in the pathways are naturally highly specific to their context, which means they are not suited to understanding the kinds of online radicalisation observed in this research. Additionally, in Australia in particular, recent years have seen an overemphasis on the vulnerability of Muslim youth to extremism, resulting in the framing of these youth as "becoming terrorists", which both stigmatises these young people and limits new research directions (Abdel-Fattah, 2020). Our proposed pathway aims to be more general in nature, tracing the behaviours, mental states, and triggers that can induce and advance online radicalisation. In particular, this model is based on research looking at former members of far-right, QAnon, and anti-vaxxer communities, who mostly departed from their radicalised communities before progressing to the extreme final stage. While some did reach this stage, we note that this is due to our expansion of the idea of what constitutes an extreme act beyond an act of violent terrorism, as will be detailed below. We suggest that this pathway could be used to understand and humanise radicalised individuals,

evaluate how far they have been drawn into fringe views, and determine appropriate interventions.

Method

Development of the pathway

This research stems from online observation of fringe groups across 2022-2023. During this observation period, we identified Facebook Groups and Pages where misinformation was consistently posted and reviewed them on an ongoing basis. We took comprehensive field notes on post content, style, and motivations for involvement in the misinformation in the community. This initial observation process formed the basis of the first draft of the pathway, which was further enriched by a second wave of research. This second stage included a close study of 27 news reports in which former believers of misinformation and conspiracy theories spoke about their journeys, for which university ethics approval was granted [REDACTED FOR REVIEW]. These were coded using NVivo software to identify behaviours and views in accordance with stages one-to-six. Additionally, we sourced 20 posts from online, organic deradicalisation spaces on social media platforms, where former and doubting believers posted about their journeys and efforts to leave their fringe communities. These posts were de-identified and also coded in NVivo in accordance with stages one-to-six. We mapped this original research against models found through the literature search discussed below to develop our pathway. However, our ethics approval was on the condition that in discussion of the data from organic deradicalization spaces, no group or individuals would be mentioned, and no direct quotes would be used, to prevent identification of users. Therefore, we are unable to provide further details about the sources beyond what we have already given. In

each instance where available, we instead provide examples from the published, public items that illustrate our observations.

Literature review

Our pathway model was primarily developed through online observation; however, we also undertook a literature search of existing scholarship to contextualise our research on 28 February 2023, using the [REDACTED FOR REVIEW] library catalogue. Rather than surveying the related literature as a whole, we conducted a systematic literature review that aimed specifically at finding existing models or pathways for radicalisation that could enrich our understanding of the pathway we developed based on our research. The searches were “‘radicalisation path*’ OR ‘radicalization path*’ or ‘radicalisation journey’” (32,322 results) and “‘radicalisation model’ OR ‘radicalization model’” (110 results). We included journal articles and book chapters, and excluded non-scholarly formats, such as dissertations, newspaper articles and reviews. The language was limited to English. For the former search, due to its large volume of search results, we reviewed 100 items, until the relevance of results veered away from the search focus. For the second search, all 110 results were reviewed for potential inclusion. Ultimately, 55 results were downloaded for close reading from these two searches. We considered articles for inclusion based on the following criteria:

- The articles either outlined some kind of staged process of radicalisation or closely examined a potential contributing factor in-depth.
- The article was not about a computational or simulated model.
- The model proposed held relevance for our own model; that is, its insights were applicable to the forms of radicalisation that we examined.
- The models themselves had been developed by scholars, and not a government or other organisations which could be tailored toward their own context and needs.

Of all the results surveyed, nine papers were ultimately included and are integrated throughout the discussion of each stage below. Six of these outlined a staged process for radicalisation that we identified as comparable to our own, and an additional three papers provided detailed insights into relevant forces that could enable the progression of radicalisation.

The pathway

The pathway below is drawn from anonymous data collected online and published newspaper interviews with ex-supporters of fringe views. Where applicable, we provide citations of *public* case studies to illustrate our examples.

Stage 1: Uncertainties and Questions

The preconditions that make individuals vulnerable to radicalisation have underlying commonalities, such as dissatisfaction with oneself, experiencing lifestyle pressures, lacking community, and general uncertainties about life. Kruglanski et al. (2022) identify three influences that make individuals vulnerable to radicalisation, including a need for significance in the world, for “narrative,” or a worldview that gives structure, meaning, and purpose to one’s existence; and a need for a network and community. Similarly, Klausen et al. (2015) argue that the pre-radicalisation stage is marked by disillusionment with the world, a personal crisis such as a death or drug addiction, and the start of an information-seeking process with existing or new authority figures. Vergani et al. (2020) label these as “pull factors”—personal factors and individual experience—which work in tandem with “push factors”—political, structural, or sociological forces—to make someone vulnerable to radicalisation. When one or both are inflamed, they act as the “transformative triggers”

described by Winter and Feixas (2019) that undermine an individual's self-concept and worldviews.

These personal vulnerabilities (Catherine and Louis, 2023) or sensitivities (van Eerten et al., 2019) may also co-occur with minor exposure to extreme content. Pepys et al. (2020) identify these factors, but draw attention to the necessity of “exposure to radicalising moral contexts, and the emergence of radicalising settings”, often political or religious, to trigger the radicalisation process. McCauley and Moskalenko (2008) consider personal victimisation and political grievance to both be powerful forces that can make an individual vulnerable, while El-Muhammady (2019) pairs features like social isolation and low self-worth with more personal motivations to action, such as experiencing cognitive distortions, seeking quick and easy means of creating change for the better, and high levels of narcissism. It is important to note, therefore, that while existing research has observed these character traits and experiences as factors as preconditions to radicalisation, they are not in and of themselves signs of fringe behaviour or views; but rather potential traits that may increase susceptibility.

In our data, the nature of radicalisation triggers provided some insight into the person. Among those who joined far-right communities, circumstances that contributed to radicalisation were a sense of loneliness and inadequacy due to either a lack of friends or negative experiences with girls when they were in school, along with a chance encounter with some kind of role model and community which welcomed them. Based on the information disclosed, the individuals in far-right communities were all male and typically identified their age of vulnerability as around 12-14 years of age when they reflected on their experiences as adults. Because this cohort was all-male, and the radicalising circumstances described often related to (in)experience with girls, we suggest the particular accounts here may not be as applicable to women in far-right communities. Only two individuals in our sample stated

their cultural identity, and both self-identified as “white” and had been part of far-right communities. Among anti-vaxxers, which included male and female users, experiences ranged from distrust in their country’s healthcare system exacerbated by the COVID-19 Pandemic to the birth of a new child and anxiety about the child experiencing pain (see, for example, Martin, 2021; Hamby et al., 2021; Schreiber, 2022; Rose, 2021; Stewart 2021,1; Idlebrook, 2021). Additional reasons included unknowingly joining a fringe community when seeking out parenting advice, and job insecurity. Among the groups examined in the development of this model, it was boys and young men who joined the far-right that had the most similar stories which depended heavily on personal characteristics; while radicalisation for anti-vaxxers typically required a stronger catalyst to trigger their adoption of fringe views.

A range of topics brought people to the point of believing in QAnon conspiracy; in a way that is perhaps reflective of QAnon’s nature as an umbrella for a range of smaller conspiracies. These included the individual having a pre-existing interest in American politics and controversies like the Snowden leaks in 2013 or struggling with mental illness, which several posters felt had made them more vulnerable to becoming believers (Carrier, 2021; Rosen, 2021; Dickson, 2021; Boigon, 2021; O’Sullivan, 2021). Several former QAnon believers identified themselves as former die-hard supporters of Bernie Sanders due to their strong distrust of governments—an element also present in the QAnon tale (Carrier, 2021). These individuals had often been sent social media content, such as YouTube videos about the QAnon conspiracy theory, by friends who shared their views, which triggered their radicalisation. Many more identified themselves as having already held right-wing or conservative views, often due to their family’s influence, and found these views were stoked by Fox News, the rhetoric of Donald Trump, and social media which gave them increasingly fringe content based on recommendation algorithms (O’Sullivan, 2021). Also unique to QAnon were identifications of religion, usually Christianity, as a channel that made one

susceptible to radicalisation due to the biblical nature of the QAnon narrative (O’Sullivan, 2021). Common to all posters was discontent with the state of the world or powerlessness in their own lives, as well as an inciting incident that caused them to seek out or come into contact with a source of misinformation and fringe views.

At this early stage of uncertainty, individuals may commonly exit the radicalisation pathway through access to reassuring advice from a trusted source. This could include information from a reliable source such as a news report, a reputable website such as a government fact sheet about vaccines, or consultation with authoritative figures such as doctors, community leaders, or even parents. It is crucial to emphasise that many people, especially people who spend a lot of time online, are routinely exposed to misinformation and experience uncertainties about topics that can fuel extreme views (El-Muhammady, 2019). The capacity to resist misinformation and fringe views “is particularly true for those who have a strong value system, self-identity and intellectual capabilities”; traits that starkly oppose those above-discussed personal characteristics that can leave one vulnerable (El-Muhammady, 2019). Nevertheless, many people can move into this first stage of radicalisation, particularly amidst periods of societal disorder, such as the COVID-19 Pandemic; however, most people resolve these doubts or dismiss the misinformation and do not progress further down this radicalisation pathway (El-Muhammady, 2019). Plain-language versions of these resources, promoted through accessible forums like social media and government websites, could reach more people in this early stage.

Stage 2: Curiosity and Criticism

Analysis of posts and interviews revealed a stage of information-seeking behaviour to clarify participants’ thoughts on their concern. For example, questioning the effectiveness of vaccinations to one’s doctor or sharing tentative disclosures of interest in fringe views to

existing social networks. In these accounts, experts and peers alike who reacted with a stigmatising reaction to the topic, like disbelief at the individual's views, or some kind of mockery or shaming, accelerated the radicalisation process for the vulnerable individual, who made a rapid retreat to a "safe" online space related to the topic for assurance.

This disclosure/stigmatisation stage did not exist in the models we investigated, possibly because many deal specifically with radicalisation related to violent extremism. As a result, those individuals would be aware that their newfound perspectives would likely not be welcomed by many people around them, and more cagily guard their activities. In contrast, doubts or confusion about medicine, politics and other government activities, or daily uncertainties are often within the scope of ordinary conversation for many people. As tentatively raising one's views on these topics is not socially egregious or inherently criminal the way violent terrorism is, those who are drawn into conspiracy theories and misinformation may be more likely to experience this stage of the radicalisation process.

In our original data, doctors, friends, and romantic partners were among the most-mentioned groups to whom vulnerable individuals disclosed their emerging fringe views. This disclosure was often a significant investment of trust for the individual, who had to genuinely work up the courage to share their thoughts and questions with others. As a result, reactions of judgement, disparagement, and aggression from peers and mentor-figures produced visceral shame and embarrassment for the individual, at times making them regret having placed their trust in that person (Martin, 2021; Hamby et al., 2021; Schreiber, 2022; Rose, 2021; Carrier, 2021). They then retreated to online spaces where they either read content that further affirmed their views or, in some cases, tentatively posted about their experience, for which they received support and encouragement from others further down the rabbit hole. This experience caused them to become further radicalised.

The emotional strength of the accounts in our sample about the impact of stigmatising reactions from peers and mentors provides a clear indication of what a deradicalising off-ramp might be at this stage: an empathetic, non-judgemental reaction from the same cohort. In one example in our sample, an individual left their fringe community at this early stage due to an empathetic doctor. It is important to recognise that such peers and mentors in the sample were not described as having been mean-spirited, but rather, having not fully recognised the sensitive nature of the issue. One method to perhaps encourage more understanding responses is creating resources for people on how to talk to a friend or associate (e.g., a patient) who is expressing fringe views. These resources could guide peers and mentors to identify the underlying issue and provide the affirmation that would otherwise be sourced from fringe online spaces. It could also guide the individual to a solution for their anxiety such as a trusted resource recommended by a friend.

Stage 3: Connection

At this stage of radicalisation, the individual has started to view themselves as part of a community centred around the issue. This internalisation of a person's views on the topic starts to transform their sense of identity, so that their 'in-group' allegiance is now more closely associated with their new fringe networks. This is recognised in existing models, such as by Winter and Feixas (2019: 4), who identify "the development of a 'turning point' in his or her sense of identity with a more structured and certain view of the world" as a key part of the process. Likewise, El-Muhammady (2019: 160) recognises this middle stage as one characterised by a "deepening" of the person's new views. In our data, more time spent online in related fringe spaces was a common behaviour that enabled the deepening investment in the new community. Specifically, this often involved more reading (which they called "research") about the topic, such as the threads of a specific QAnon theory or details

about vaccine reactions. At this stage of deepening investment, many people reported still feeling grimly about the topic with which they were concerned. However, they gained an increasingly addictive thrill from “connecting the dots” between their own anxieties and plausible explanations online.

Another feature identified during this stage is the withdrawal from existing networks, and the beginnings of perceiving those with different views as so-called outsiders. This can manifest in behaviours like the self-censorship of discussion about their topic(s) of concern with those existing peers, and a withdrawal from mainstream information sources. As Winter and Feixas (2019) note in steps 5 and 6 of their model, and McCauley and Moskalenko note in stage 4 (2008) of their model, increased negative perceptions of an opposing group as well as “constriction of their previous social world” also occur around this time. Some individuals we observed reflected that as their investment in their new views intensified, they began to behave less kindly towards people they knew who were “outsiders” to their views, such as women or people with different views on medicine. Rather than casting this trio of changes as three different phases, in our model, we see these behaviours as co-occurring as the individual re-negotiates their own identity. As a result of their increased reliance on fringe communities for social and informational needs, which validates their feelings, we also include in our model the emergence of feelings of happiness, certainty, hope, and reassurance. The alleviation of the earlier anxieties that drove them to this community are partly driven by the discovery of a new community they trust, that can explain away lingering doubts and provide steps an individual can take to achieve the desired stability in their own life.

At this stage in our sample, when the individual was withdrawing from mainstream informational sources and their existing peer networks, there were fewer opportunities for traditional fact-checking methods to reach them. The value of organic de-radicalisation

spaces becomes apparent at this point (Thorburn, 2023b). These spaces allow radicalised individuals to speak to others who have previously shared their beliefs, allowing them to be understood whilst also being provided with more reliable information. In our sample, there were some instances where people stumbled on these spaces while caught in a conspiracy-related research spiral, and this triggered the realisation that they wanted to ‘get out’, but did not know how to begin. They were able to make posts to the space explaining their situation, and former believers advised on how to leave the fringe community. Many users expressed immense shame at having been involved in a fringe community and did not want to discuss it with a therapist or friend for fear of admitting what they had believed; however, former believers were able to withhold judgement and gain the trust of new arrivals. Based on the posts there, it is clear that these spaces offer a unique form of support. Finding ways to promote and support such spaces, either through targeted algorithm recommendations or other interventions, could increase their reach and effectiveness.

Another exit path that emerged at this stage was the individual limiting or leaving social media and the related online community. Individuals did this willingly, usually following the realisation that their investment in this community was worsening their insecurities rather than improving their life (Carrier, 2021). For example, noticing that outcomes promised by their online community were not fulfilled, or that their mental or physical health was declining. The decrease in online involvement, which could be spent on hobbies or with friends and partners, gave them perspective on their lives and positive experiences of the world. This allowed them to re-evaluate the impact of the online community and its views on their life, their goals, and the world around them. Given that research has illustrated that time spent online and on social media can be detrimental to human health while limiting use brings benefits (see Hartman and Quick, 2022; Hunt et al., 2018), the impact of living a more analogue life should not come as a surprise. In terms of

ways to trigger this exit path at this stage, more public awareness could be raised about the impacts of social media, and crucially, how to recognise excessive digital usage and “log off”. This could be promoted online and through school education to reach younger age groups to develop healthy digital habits at a younger age and reduce exposure to misinformation in the first place.

Stage 4: Community Investment

At this stage of deep immersion in the fringe community and its beliefs, most research identified features such as lifestyle change, change in peers, and change in worldview as key components. Winter and Feixas (2019) emphasise the increased hostility of individuals from stages 7-10 of their model, including a more derogatory view of out-group members and an increased acceptance of violence. Van Eerten et al. (2019), drawing on their use of the penultimate stage of Doosje et al.’s model, identify the adoption of group membership as crucial to the increasing depth of the person’s beliefs. At this stage, the group identity provides a kind of “shield” against de-radicalisation efforts because the radicalised individual values the community itself (Van Eerten et al., 2019: 29-30). Catherine and Louis (2023) also draw attention to the role of a group’s worldview and their dynamics in these late stages of radicalisation, highlighting the significance of the newly acquired identity in keeping the individual radicalised. Several individuals in our data revealed that efforts to talk them out of their views based on logic were ineffective. Instead, they became more adept at defending their new views when they arose in conversation with their existing social networks. Our model also notes that the individual’s reliance on the community for social needs and a sense of self has become a dependence at this stage, with individuals in our case studies reporting an increasingly black-and-white or us-vs-them mindset, and, in some instances, an increased comfort with the idea of violence in the name of their group. We particularly noted this in

relation to women and perceived political opponents, with both cast as conspirators who had to be brought under control. As a result, this animosity could emerge during encounters with their day-to-day peer network, no longer concealed from view, and this could prompt increased stigma from existing networks and a further loss of support.

We also identified several features more broadly relevant to fringe communities that largely operate online. For example, spending large amounts of time online to one's own detriment, such as causing serious sleep deprivation, illness, and a loss of social and leisure time. Klausen et al. (2015: 72) also noted this; however, given that their model is largely concerned with extremism that can lead to terrorism, the role of excessive time spent online is placed earlier in their 'detachment' phase. Klausen et al. (2015) also identify increased hostility to others and making seemingly abrupt changes to one's lifestyle as traits of this stage, which in our model were mostly identified when people felt they were most deeply immersed in their online radicalised space. The investment in this space could include some kind of active participation by this stage, such as contributing one's own interpretations of (mis)information, and sharing their own sources with others—again, a behavioural emphasis on information that is not as emphasised in other models we have discussed. For many, it was the consumption of misinformation and positioning of the self within the greater narrative they were consuming that provided a sense of agency and action, even if they were not actively preparing for violence themselves (Carrier, 2021; Dickson, 2021). However, one case study who had joined a far-right group and had an established offline community by this stage began bracing themselves for participation in violence by starting to acquire weapons, indicating that perhaps face-to-face connections can prompt one to some kind of action at an earlier stage.

As is described in Catherine and Louis (2023: 190), the new-found sense of identity, community, and source of guidance in one's life sometimes leads to feelings of euphoria, and

a strong sense of being among the “special” or “chosen” humans who had been called to a particular cause. This euphoric feeling could also be interspersed with periods of severe doubt, which were soothed by the new social network that offered reassurances and answers to these doubts. For at least one individual, the euphoria brought about by the deep involvement in QAnon supported them through periods of severe depression related to their offline life circumstances. The role of the community as an information source had the dual effect of maintaining an individual’s social ties to the group, while restricting any outside input of information which could otherwise challenge their worldview (Dickson, 2021). At least one user even actively sought to prevent members of their online community from disengaging from the community and abandoning its views, thus preserving the community and worldview. In this insulated state, efforts to reach the radicalised individual through fact-based discussion also had to contend with the individual’s new sense of identity and community ties, which provided a strong defence against outside influence.

We also include the individual’s active withdrawal from existing social networks as another key manifestation of this stage. While the person may have already distanced themselves from weaker social links, at this later stage, we found that they made proactive decisions to sever ties with friends and family. This could be due to the belief that the person was holding them back, untrustworthy, or even somehow dangerous to the radicalised individual. These decisions caused much grief for those who were cut off, according to former radicalised individuals who were now aiming to repair those connections.

At this stage, some previously discussed exits from the radicalisation pathway still benefited individuals, such as a trusted, empathetic friend, limiting social media use, or organic deradicalisation spaces. However, at this higher level of investment in the fringe community and active withdrawal from existing networks, we also saw the emergence of several new exit pathways. The first of two new forms was the experience of a personal

shock that forced them to reconsider their views. This could be the death of a family member or friend, a significant personal realisation (e.g., sexual orientation), or something related to the fringe community itself, such as election outcomes (Stewart, 2021; O’Sullivan, 2021; Idlebrook, 2021). These experiences were so deeply personal that the individual was forced to confront their views to grapple with the new circumstances of their life. As a result of this need to reconcile their new reality with the alternative reality provided by their fringe community, many of the individuals in our sample were able to begin withdrawing from those spaces and make efforts to rebuild their lives. This indicates that intervention methods at this stage need to be able to reach an individual at a deeply personal level, rather than relying on facts, debate, and logic.

In contrast, the second new exit pathway that emerged at this stage was triggered by a source that was detached from the individual. Several people in our sample were prompted to re-evaluate their views after being exposed to third-party content that exposed the falsehoods of a fringe view (Carrier, 2021; Dickson, 2021). This could come in the form of authoritative documentaries about other conspiracies or scams or user-generated content that impartially highlighted the illogic of a fringe view. Crucially, none of these were government-made; rather, what seemed to make the content effective was that it was by a party not trying to debate them, but simply presenting facts in a clear, unbiased way. This raises possibilities around whether recommendation algorithms on digital media platforms could be harnessed to help promote this content, as well as making relevant documentaries more available. Uniquely, the value of this content was its impartial approach to revealing how conspiracies, scams, and misinformation can be created and promoted, as well as the fact that it did *not* come from an authoritative source. While direct fact-checking merely made people feel under attack and defensive, individuals could be reached with this content as they had left their guards down.

Stage 5: Erratic behaviour

At stage five, existing models and our own model show clearer signs of divergence. In common with ours is the view that, at a second-to-last stage, individuals are feeling an increasing sense of need to ‘give back’ to the group they are a part of, or ‘take action on behalf’ of the group’s goals. This stage of heightened motivation and even preparation is understood by Klausen et al. (2015: 72) to be a stage when one involved in religious radicalisation may undergo training with peers, to learn how to carry out violent extremist attacks. For El-Muhammady (2019: 174), this stage is similarly defined by individuals “externalising” their views, which might include working together with others. At this stage, the individual’s radicalisation is now visible to others as more than a passing eccentricity. Physical manifestations of erratic views in the individual’s lifestyle are more broadly defined in our model, to reflect the wider range of radicalisation experiences we observed. For example, rather than needing to flee overseas for wartime training, the internet allows those seeking relevant skills and materials to source them online. Furthermore, the existence of digital platforms allows for potential participation in public displays of social division that do not require training or violence. For example, participating in hashtag campaigns to trend their issues in mainstream social media spaces potentially recruiting new members. Likewise, in another example in our dataset, an individual performed a targeted act of vandalism, a social disruption that did not cause physical harm to another person but was an expression of their views.

At this stage, the two exit paths in our sample were highly similar in nature. The first was medical intervention, often brought about by an existing issue, such as for anxiety, depression, ADHD, or addiction. Posters did not give context for whether they sought this help themselves or whether they were urged to do so by people around them. However, they did reflect on how addressing the severe personal health issues they were dealing with had

triggered their deradicalisation process. As their wellbeing stabilised, they could take stock of their life and battle their dependency on coping mechanisms, including the fringe communities and worldviews. While not every radicalised person necessarily has health problems, or the means to seek professional support for health problems, this does indicate an alternative route by which authoritative sources such as medical professionals can enter the lives of these individuals as a trusted source of support.

The second way deradicalisation was triggered at this stage was by being forcibly institutionalised for mental health and related issues, after a severe breakdown (Boigon, 2021). Again, while this is not necessarily applicable to all radicalised people, it reaffirms the level of effort and resources required to reach someone at this stage. From these examples, we can observe the importance of providing a new structure or plan for the future for the individual, to help with the loss of the fringe community and views they depended upon. Creating clearer channels accessible to the public, such as online resources and support groups, to help families help loved ones leave these communities could provide more opportunities for stronger interventions.

Stage 6: Committing Harm

Overt extremist acts were uncommon in the data we examined, likely due to the posters having parted ways with their fringe community and many of the associated views. However, numerous examples of people who have gone on to commit these acts are available in news media, ranging from QAnon supporters storming the US Capitol building (Barry et al., 2021), to more individualised actions such as deciding not to vaccinate one's child. Our model offers a broader recognition of extreme acts in line with the types of radicalisation observed in our sample and in the online fringe spaces we monitored. We acknowledge that there are potential acts that users could commit (or have committed) that they would choose not to

disclose on an online forum. Therefore, a stage six extreme act in our model could be anything that may negatively and *materially* impact someone, including solely themselves, or others; and which is completed with intention and ideological belief.

In contrast, in the models considered in the existing literature, there is a common understanding that the final stage of the radicalisation process entails a violent extremist act, such as a suicide bombing. The focus on violent acts in existing models is likely due to the original scope of these works on violent and religious extremism. While the models themselves are appropriate to these scenarios, they cannot be used to fully understand the newer kinds of radicalised acts that are taking place in an age where misinformation and conspiracy are so easily transmitted across digital spaces and apply to so many topics. For example, in our dataset, the most extreme self-confessed act was among anti-vaxxers, who decided to stop vaccinating their children (Martin, 2021; Hamby et al., 2021; Schreiber, 2022; Rose, 2021; Stewart, 2021; Idldebrook, 2021). Given the considerable threat this poses to a young child's life even under normal circumstances, to say nothing of the risk posed by COVID-19, this decision constitutes an extreme act within the context of an anti-vaxxer community. In the context of other communities in our dataset, which were the far-right and QAnon, there was a greater appetite for violence revealed in earlier stages, but no self-disclosed acts in these views. This pattern supports Kruglanski et al.'s observation that most radicalised individuals do not go on to commit extreme acts (2022). Rather, in the framing of our model, these individuals would hover somewhere in the middle stages of the model—stages four or five—where they fuel the online culture of their fringe community, but do not act on their fringe views in a manner that causes direct material harm. At stage six, for those embedded in a fringe community with an appetite for violence, the primary exit path would likely be a deradicalisation program such as one organised by a government organisation, as has been demonstrated in previous academic studies (El-Muhammady, 2019). However, our

sample showed no such examples of this extreme final stage among the more violent QAnon or incel communities.

Conclusion

This article presents a qualitative model of different stages of radicalisation that individuals may progress through, which is suitable to contemporary forms of online misinformation and conspiracy theories. It draws on a combination of existing scholarship, as well as analysis of case studies of former believers through interviews with newspaper publications and of 20 anonymised posts from online de-radicalisation spaces. From this dynamic combination of sources, we presented a portrait of the emotional and behavioural experiences one might have at various stages in the pathway should their fringe views increase in intensity. We also identify what interventions could be most impactful at each stage. We emphasise that while the majority of people who are radicalised into extreme communities do not reach the later stages where they may commit acts of violence, those in the early and middle stages are still a cause for concern as their actions fuel the misinformation ecosystem that eventually generates violent criminals. This threat will likely grow in the future, as exposure to misinformation, conspiracy theories, and other radicalising content is increasingly easy for people to experience, given the ubiquity of digital communication methods.

Misinformation radicalisation is a complex problem because, at its heart, it is a very human problem. In the data analysed for this paper, it was largely underlying personal uncertainties about life and specific topics that made people vulnerable to adopting fringe views. This is to say that while the issue of misinformation necessarily involves a topic about which one may adopt fringe views, people are not radicalised *because* of information. As a result, when it comes to protecting people from being drawn into fringe communities, we need to think beyond fact-checking and address the underlying social issues that drove people

to adopt these views: distrust in governments and other figures of authority such as doctors, mental illnesses, and a culture that pressures young men and boys to measure their success in life by their acquisition of a girlfriend. Such change needs to be addressed from many angles, ranging from increased government funding for mental health services and outreach to ensure these services are used, including in schools. Greater visibility of men with a healthy relationship with women and themselves, and of approachable advocates for medicine and personal wellbeing, could also be beneficial and accomplished through partnerships with respected public figures. Uncertainty is a natural part of the human condition that cannot be eradicated, but accessible health services and trustworthy role models are just some of the ways we can help people safely navigate these feelings.

We also recommend smaller-scale actions that could be undertaken in the near future. The development and distribution of accessible, engaging resources about topics of concern by governments or affiliated bodies could provide the reassurance many people need when fears arise concerning a particular issue. Likewise, we recommend the development of resources to assist professionals and members of the public in speaking to a person who may be beginning to express fringe views, to maintain the relationship and keep the person connected to the outside world. The promotion of organic de-radicalisation spaces online, or partnerships with nonprofits that seek to help people escape fringe communities such as Life After Hate (<https://www.lifeafterhate.org/>), Back to the Vax (<https://backtothevax.com/>), and Freedom of Mind (<https://freedomofmind.com/>) could also serve as a lifeline for these vulnerable people. The promotion of such spaces can be achieved by tweaking the infamous social platform algorithms, which have contributed to driving their users down radicalisation rabbit holes as found in our sample. We also suggest that, even if could be challenging to implement, the companies behind these platforms might need to be held responsible for the effects their automated tools have on our society. Furthermore, encouraging people to limit

time spent online through public awareness campaigns about social media's detrimental impacts could also benefit both adults and young people. These could be government-led or integrated into existing health and wellbeing programs in schools and workplaces. Lastly, we emphasise that everyone can contribute to addressing the issue by responding with empathy if they discover someone they know has a newfound interest in fringe views. Understanding the misinformation radicalisation process as a pathway driven by human emotions rather than the information itself is the first step in humanising these individuals.

References

Abdel-Fattah, R. (2020). 'Countering violent extremism, governmentality and Australian Muslim youth as "becoming terrorist"', *Journal of Sociology* 56(3): 372-387.

<https://doi.org/10.1177/1440783319842666>

Attwell, K., Smith, D. T., and Ward, P. R. (2021). "'If your child's vaccinated, why do you care about mine?' Rhetoric, responsibility, power and vaccine rejection', *Journal of Sociology* 57(2): 268-285.

<https://doi.org/10.1177/1440783319893468>

Barry, D., McIntire, M., and Rosenberg, M. (2021). "'Our President Wants Us Here': The mob that stormed the Capitol', *The New York Times*, 9 January, URL (consulted 3 August 2023):

<https://www.nytimes.com/2021/01/09/us/capitol-rioters.html>

Boigon, M. (2021). 'She was a Jewish QAnon supporter. And she warns it could happen to you.',

Haaretz.com, 11 February, URL (consulted 3 August 2023): <https://www.haaretz.com/us-news/2021-02-11/ty-article/she-was-a-jewish-qanon-supporter-and-she-warns-it-could-happen-to-you/0000017f-e566-d62c-a1ff-fd7f7f0f0000>

Carrier, A. (2021). 'QAnon Almost Destroyed My Relationship. Then My Relationship Saved Me From QAnon', *Politico*, 13 August, URL (consulted 3 August 2023):

<https://www.politico.com/news/magazine/2021/08/13/qanon-radicalization-bernie-sanders-supporter-503295>

- Catherine, B., and Louis, B. (2023). 'A Path of Radicalization: Complementary of the Group and the Individual Flaw', *Journal of Forensic Psychology Research and Practice* 23(2): 178-200, <https://doi.org/10.1080/24732850.2021.2016123>
- Dahlke, R., and Hancock, J. (2023). 'The Effect of Online Misinformation Exposure on False Election Beliefs' [Preprint], OSFPreprints: 1-23. <https://doi.org/10.31219/osf.io/325tn>
- Dickson, E. (2020). 'How Ex-QAnon Followers Escaped The Cultish Conspiracy Theory', *Rolling Stone*, 23 September, URL (consulted 3 August 2023): <https://www.rollingstone.com/culture/culture-features/ex-qanon-followers-cult-conspiracy-theory-pizzagate-1064076/>
- Eerten, J.-J. v., Doosje, B., Konijn, E., and Graaf, B. d. (2019). 'Counter-Messaging: An effective means to prevent or counter radicalization?', pp. 27-44 in J.-J. v. Eerten and B. Doosje (eds) *Challenging Extremist Views on Social Media*. London: Routledge.
- El-Muhammady, A. (2019). 'Radicalisation model: Learning from Malaysian militant-extremists' pp. 155-184 in R. Gunaratna and S. Hussin (eds), *Terrorist Deradicalisation in Global Contexts*. London: Routledge.
- Evelyn, K. (2020). 'Advocates blame anti-vaxxers after four-year-old boy dies from flu', *The Guardian*, 3 August, URL (consulted 3 August 2023): <https://www.theguardian.com/us-news/2020/feb/07/colorado-boy-dies-flu-anti-vaccine-facebook-groups>
- Gerrand, V. (2022) 'Communicative channels for pro-social resilience in an age of polarization', *First Monday* 27(5). <https://dx.doi.org/10.5210/fm.v27i5.12599>
- Gillespie, E., and McGowan, M. (2022). 'Queensland shooting: Gareth and Stacey Train published YouTube video after killing police officers', *The Guardian*, 26 December, URL (consulted 3 August 2023): <https://www.theguardian.com/australia-news/2022/dec/16/queensland-shooting-gareth-and-stacey-train-youtube-video-published-after-killing-police>
- Green, R., Fowler, K., and Palombi, A. (2023). 'The Black pill pipeline: A process-tracing analysis of the Incel's continuum of violent radicalization', *Sexualities* 0(0): 1-21. <https://doi.org/10.1177/13634607231169509>

Hamby, B., Mullis, S., and Martin, R. (2021). 'She Resisted Getting Her Kids The Usual Vaccines. Then The Pandemic Hit', *NPR*, 22 January, URL (consulted 3 August 2023):

<https://www.npr.org/2021/01/22/956935520/she-resisted-getting-her-kids-the-usual-vaccines-then-the-pandemic-hit>

Hartman, D. E., and Quick, B. L. (2022). 'A Reasoned Action Approach to Limiting Excessive Social Media Usage Among Adults', *Health Communication*: 1-10.

<https://doi.org/10.1080/10410236.2022.2129315>

Hunt, M. G., Marx, R., Lipson, C., and Young, J. (2018). 'No More FOMO: Limiting Social Media Decreases Loneliness And Depression', *Journal of Social and Clinical Psychology* 37(10): 751-768.

Idlebrook, C. (2021). 'I was once a hardcore anti-vaxxer. Now I try to nudge people to get the Covid-19 vaccine', *Start News*, 29 August, URL (consulted 3 August 2023):

<https://www.statnews.com/2021/08/29/former-anti-vaxxer-now-nudge-people-to-get-covid-19-vaccine/>

Jytte Klausen, S. C., Nathan Needle, Giang Nguyen and Rosanne Libretti. (2015). 'Toward a Behavioral Model of "Homegrown" Radicalization Trajectories', *Studies in Conflict and Terrorism* 39(1): 67-83. <https://doi.org/10.1080/1057610X.2015.1099995>

Kruglanski, A. W., Molinario, E., Ellenberg, M., and Cicco, G. D. (2022). 'Terrorism and conspiracy theories: A view from the 3N model of radicalization', *Current Opinion in Psychology* 47: 1-5.

<https://doi.org/doi.org/10.1016/j.copsyc.2022.101396>

Martin, R. (2021). 'Former Anti-Vaccine Mom Explains How Movement Pulled Her In, And How She Left', *NPR*, 18 January, URL (consulted 3 August 2023):

<https://www.npr.org/2021/01/18/957981974/how-anti-vaccine-movement-could-hurt-efforts-to-end-pandemic>

McCauley, C., and Moskalenko, S. (2008). 'Mechanisms of Political Radicalization: Pathways Toward Terrorism', *Terrorism and Political Violence* 20(3): 415-433.

<https://doi.org/10.1080/09546550802073367>

O'Sullivan, D. (2021). 'She was stunned by Biden's inauguration. How this South Carolina mom escaped QAnon', *CNN*, 3 February, URL (consulted 3 August 2023):

<https://edition.cnn.com/2021/02/03/tech/qanon-mom-former-believer/index.html>

Pepys, R., Bowles, R., and Bouhana, N. (2020). 'A Simulation Model of the Radicalisation Process Based on the IVEE Theoretical Framework', *Journal of Artificial Societies and Social Simulation* 23(3): 1-22. <https://doi.org/10.18564/jasss.4345>

Rose, A. (2021). 'Former Antvaxxer Reveals Why She Got the Covid Vaccine', *News.com*, 20 July, URL (consulted 3 August 2023): <https://www.news.com.au/lifestyle/health/health-problems/former-antivaxxer-reveals-why-she-got-the-covid-vaccine/news-story/aea124bd60fc9988e291ae4ef40186e6>

Rosen, J. (2021). 'Former QAnon believer says following the conspiracy "was absolutely a drug"', *CBS News*, 14 February, URL (consulted 3 August 2023): <https://www.cbsnews.com/news/qanon-conspiracy-believer-drug/>

Schreiber, M. (2022). "'The only logical choice": anti-vaxxers who changed their minds on Covid vaccines', *The Guardian*, 14 February, URL (consulted 3 August 2023):

<https://www.theguardian.com/us-news/2022/feb/14/covid-anti-vaxxers-changed-minds>

Scrivens, R., Venkatesh, V., Bérubé, M., and Gaudette, T. (2022) Combating Violent Extremism: Voices of Former Right-Wing Extremists, *Studies in Conflict & Terrorism*. 45:8, 661-681.

<https://doi.org/10.1080/1057610X.2019.1686856>

Stewart, C. (2021). 'Former anti-vaccine influencer now using social media to encourage others to get vaccinated', *Denver7*, 15 October, URL (consulted 3 August 2023):

<https://www.denver7.com/news/national-politics/the-race/former-anti-vaccine-influencer-now-using-social-media-to-encourage-others-to-get-vaccinated>

Thorburn, J. (2023a). 'Exiting the Manosphere. A Gendered Analysis of Radicalization, Diversion and Deradicalization Narratives from r/IncelExit and r/ExRedPill', *Studies in Conflict & Terrorism*.

<https://doi.org/10.1080/1057610X.2023.2244192>

Thorburn, J. (2023b). 'The (de-)radical(-ising) potential of r/incelexit and r/exredpill', *European Journal of Cultural Studies* 26(3): 464-471.

<https://doi.org/https://doi.org/10.1177/1367549423115390>

Vergani, M., Barton, G., and Iqbal, M. (2017). 'Beyond social relationships: Investigating positive and negative attitudes towards violent protest within the same social movement', *Journal of Sociology* 53(2): 445-460. <https://doi.org/10.1177/1440783316688344>

Vergani, M., Iqbal, M., Ilbahar, E., and Barton, G. (2020). 'The Three Ps of Radicalization: Push, Pull and Personal. A Systematic Scoping Review of the Scientific Evidence about Radicalization Into Violent Extremism', *Studies in Conflict and Terrorism* 43(10): 854–885.
<https://doi.org/10.1080/1057610X.2018.1505686>

Winter, D. A., and Feixas, G. (2019). 'Toward a Constructivist Model of Radicalization and Deradicalization: A Conceptual and Methodological Proposal', *Frontiers in Psychology* 10(412): 1-11. <https://doi.org/10.3389/fpsyg.2019.00412>