Current Psychology (2026) 45:358
https://doi.org/10.1007/s12144-025-08944-z

The radicalisation loop: A layered linguistic model of disengagement and re-engagement in an incel forum

Melissa S. de Roos[1] · **Giorgia Caon**[1]

Received: 25 August 2025 / Accepted: 10 December 2025 / Published online: 28 January 2026

Abstract

Online engagement with extremist communities presents significant mental health and real-world risks. This study investigated how patterns of engagement within an incel forum relate to dynamic affective and cognitive linguistic markers. Using longitudinal linear mixed models and Linguistic Inquiry and Word Count (LIWC) on a dataset of 2,009,875 comments, spanning six years, we compared users who permanently left (Desisters), those who persisted (Persisters), and those who left but later returned (Returners). Whilst overall differences between groups were minimal, distinct psychological trajectories emerged: Desisters showed an increase in grievances and hostility over time, whereas Persisters showed more stable emotions and increased social clout. Returners showed lower desperation and frustration following a period of disengagement, yet their re-engagement was marked by a significant decline in analytic thinking and a persistence of hate, signalling re-radicalisation. These findings highlight a novel, layered model of online radicalisation, offering insights for developing targeted clinical interventions.

Keywords Online radicalisation · Mental health · Emotions · Cognitive reasoning · Incels · Linguistic analysis · Mixed effects

The growing use of online platforms and social media has brought with it concerning trends, including the rise of online communities centred around hate. Recent findings suggest that 48% of users on popular platforms such as Facebook, X, and Reddit are exposed to some form of extremist content or hate speech (Williams et al., 2023). Of even greater concern are fringe platforms like 8chan, 4chan, and Gab, which are regarded as high-risk environments due to the substantial presence of extremist material on these platforms (Waldek et al., 2020). Extremist communities on these platforms are often politically motivated, especially within alt-right circles, and commonly disseminate conspiracy theories and xenophobic, racist, and homophobic rhetoric (Rieger et al., 2021). Much of the harmful content shared on these sites remains unchecked, largely because of their strong advocacy for unrestricted freedom of speech. Moreover, with the addition of fast-paced resharing and reposting behaviour of members, most of such content is not removed or challenged (Jardine, 2019).

Exposure to online hate and extremist content has been shown to negatively affect individuals' mental health and behaviour, regardless of whether the exposure is intentional (Madriaza et al., 2025). Such exposure has been linked to increased levels of depression, anxiety, stress and even emotional trauma (Madriaza et al., 2025). In terms of interpersonal functioning, online engagement with hateful extremist content has been associated with reduced social trust and the intensification of discriminatory and stereotypical attitudes (Castaño-Pulgarín et al., 2021; Näsi et al., 2015). Beyond harmful consequences for the individual, prolonged exposure to and participation in these communities further deepen the endorsement of extremist beliefs, both online and offline (Pauwels & Schils, 2016). Exposure to online hate impacts individuals by reinforcing their hateful attitudes and beliefs and is directly linked to increased engagement in online hate speech, as well as offline violent behaviour (Hassan et al., 2018; Madriaza et al., 2025). As demonstrated, online hate and extremist content have a severe impact on mental health, attitudes, and behaviour. These effects are aggravated by

✉ Melissa S. de Roos
deroos@essb.eur.nl

[1] Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

active involvement in extremist communities, where constant exposure to large volumes of hateful and violent content intensifies radicalisation and psychological distress.

Through a dynamic group effect, online communities may form echo chambers, environments in which individuals are continuously reinforced for expressing a specific point of view, while opposing views are systematically ignored (e.g. Liu et al., 2025). Through this process of selective engagement, individuals become more certain about the veracity of their opinions, ultimately even leading them to take more extreme positions over time (De Roos et al., 2024). This phenomenon is also known as group polarisation, and it inherently involves a decline in analytical, critical thinking in favour of reinforcing a group's shared ideology (Brugnoli et al., 2019). In online extremist communities, the internet appears to facilitate and accelerate these dynamics. Individuals are both passively exposed to extremist and ideologically similar content and are also actively involved in sharing and validating each other's views (Hartmann et al., 2025). Combined with the anonymity and perceived safety provided by the online platforms, such environments function as echo chambers that influence individuals through legitimising, strengthening and shaping their ideology, as well as reinforcing the sense of community (Mølmen & Ravndal, 2021).

The effects of these echo chambers extend beyond the space of extremist communities, as members increasingly use violent and hateful rhetoric in mainstream online spaces (Schmitz et al., 2024), with some people then migrating towards extremist communities. More alarmingly, some members of online extremist communities, continuously exposed to hateful narratives towards out-groups, have gone on to commit acts of real-world violence (Castaño-Pulgarín et al., 2021). Yet, despite these documented risks and the violence they can inspire, online extremist communities may also provide a sense of social connection and support for their members, particularly for individuals experiencing social isolation (Gaudette et al., 2020).

Individuals who are radicalised online often have a history of social isolation, making them more vulnerable to the appeal of communities that offer a strong sense of belonging (de Roy van Zuijdewijn & Bakker, 2016). In such communities, a strong sense of belonging is mainly built by emphasising in-group versus out-group differences, often through the collective derogation of a shared "enemy" (Walther, 2022). This helps in reducing the feeling of isolation, due to the strong sense of community. Moreover, these communities provide a space for members to share personal grievances and experiences, fostering a form of emotional and social support despite the hateful nature of the community (Døving & Emberland, 2021). Over time, in-group identification tends to strengthen, as shown in a recent study of a banned Reddit community where the sentiment around "I" and "we" became increasingly positive, while sentiment around "them" became more negative (Dilkes, 2023). Overall, for individuals feeling marginalised and misunderstood, these communities provide acceptance and validation of their opinions (Leite-Mendonca & Boislard, 2024). This creates a vicious cycle of the group's reinforcement of extremist attitudes, while reducing the sensitivity to broader societal feedback (Brady et al., 2017).

Although online communities provide a sense of belonging, they also tend to enforce social hierarchies. Research indicates that members often establish social status and influence, or clout, by using language that reinforces group norms, validates shared beliefs and demonstrates their ideological commitment (Baele et al., 2022). This dynamic can be particularly powerful in echo chambers, where the continuous exposure to like-minded views reinforces a shared ideology and rewards those who express it in an authentic and compelling way. The use of language that signals this ideological alignment and social standing can be an indicator of a user's deepening involvement. Analysing these linguistic markers, in addition to emotional content, is essential to understanding the complex dynamics of user engagement.

Individuals' susceptibility to online extremist communities can be explained by an interplay of personality traits and external or situational factors. Those who already hold internalised beliefs aligned with extremist ideologies are likely to actively engage with the community following initial exposure (Reeve, 2019). However, radicalisation can also occur through online information seeking without pre-existing extremist views (Chua, 2025). In this context, radicalisation refers to the process of adopting extremist beliefs or attitudes, whereas violent extremism denotes the behavioural expression of these beliefs through the endorsement or use of violence. Certain traits, such as impulsiveness (King & Taylor, 2011) and increased negative emotions like fear and anger (Wolfowicz et al., 2021) also increase individuals' susceptibility to radicalisation within extremist communities. These emotions often stem from perceived injustice, marginalisation, and inequality, which are common themes in extremist rhetoric (Costello et al., 2018; Schils & Verhage, 2017). Dissatisfaction with a political or societal situation may also drive individuals toward these communities to express grievances (Costello et al., 2018). However, the feelings of injustice and marginalisation can also be developed in response to broader societal events in which the individual was not directly involved (Schils & Verhage, 2017). Additionally, engagement with extremist communities is frequently driven by a desire for belonging, which these groups foster through echo chambers that continually reinforce shared views and opinions (Schils & Verhage, 2017).

Although limited, research into disengagement from both online and offline extremist communities has highlighted several common factors driving the decision-making. One important factor is disillusionment, that is, when individuals realise a dramatic difference between their expectations from the community and its actual conduct (Simi et al., 2019). This can lead to feelings of anger, shame, and even self-directed frustration (Simi et al., 2017, 2019), sometimes described as "movement burnout" (Gaudette et al., 2022). Improvement of individuals' social life, including forming meaningful relationships with family, peers or romantic partners has also been identified as a key factor in disengagement (Simi et al., 2019). This aligns with findings that social isolation and loneliness are significant risk factors for the development of hateful and extremist attitudes (de Roy van Zuijdewijn & Bakker, 2016), given that access to a broader and more supportive social network can reduce feelings of isolation and ultimately diminish the appeal of extremist communities (Simi et al., 2019). As noted by Simi et al. (2017), the disengagement process is similar to overcoming addiction. Exposure to some form of extremist content, or encounters with former community members, even long after disengagement, might cause psychological and behavioural responses rooted in individual's former extremist identity. Thus, individuals once involved in extremist communities still retain vulnerability towards rejoining such groups. However, the process and factors driving the process of re-engagement requires further investigation.

Incels as a case study in online extremism

The incel community is one such group that exemplifies the characteristics commonly associated with extremist movements. The incel community, with its established patterns of hostile rhetoric and documented links to mass violence, provides a critically salient case study. This unique context allows us to bridge the gap between general theories of online radicalisation and the empirical measurement of linguistic risk escalation over time. In fact, in the recent decade there has been an increase in acts of mass violence linked to incels, with notable incidents including the 2014 Isla Vista attack and the 2018 Toronto van attack (Tastenhoye et al., 2023). The term incel, short for "involuntary celibate," is the label adopted by individuals who perceive themselves as unable to have sexual or romantic relationships with women despite their desire to do so (Broyd et al., 2022). Incels believe that society operates according to a rigid hierarchy in which women choose partners based primarily on physical attractiveness, wealth, and social status (Fontanesi et al., 2024). From this perspective, young and attractive women are seen as highly selective, favouring men at the top of the social ladder and leaving those perceived to be at the bottom with few or no romantic prospects (Fontanesi et al., 2024).

A key feature of the incel culture is that it exists almost entirely online, with an estimated 100,000 members that continues to expand, spread across multiple platforms such as Reddit and 4chan (Daly & Reed, 2022). Generally, those belonging to the incel community express feelings of frustration and loneliness which are reinforced and validated by a rhetoric that blames women for their perceived marginalisation from society (Andersen, 2024). This externalisation of blame, which is a core tenet of incel ideology, can provide a form of temporary psychological relief by shifting responsibility for their distress from the self to an external group. However, the very same process also reinforces a worldview that perpetuates these grievances. Thus, these virtual environments offer members a space to express their grievances and receive a degree of support, which can strengthen their sense of belonging within the community (Speckhard et al., 2021).

Various shared grievances, along with a tendency toward interpersonal victimhood, often drive individuals to seek out incel communities (De Vettor et al., 2025; Fontanesi et al., 2024). Research frequently highlights common life experiences (De Vettor et al., 2025; Speckhard et al., 2021) and psychological challenges (Delaney et al., 2024; Fontanesi et al., 2024) that appear to underpin the pathways leading individuals to join these groups. Indeed, thematic analyses of content shared in incel forums reveal that many users express feelings of severe loneliness, often against the backdrop of problematic family dynamics and even, in some cases, domestic violence victimisation (De Vettor et al., 2025). Additional research shows that the majority of self-identifying incels recall feeling socially excluded during their school years, frequently citing romantic rejection and other perceived emasculating experiences, such as poor athletic ability, as formative events that significantly impacted their self-image (Speckhard et al., 2021). Moreover, studies show that many self-identifying incels report experiencing bullying throughout their childhood and adolescent years (Moskalenko et al., 2022a, 2022b) and they report bullying at a higher rate than their non-incel peers (Costello et al., 2025). In terms of psychological challenges, studies consistently report elevated rates of depression and anxiety among incel populations compared to age- and gender-matched peers in the general population (Delaney et al., 2024; Fontanesi et al., 2024; Speckhard et al., 2021). Research also highlights a high prevalence of self-reported autism-related traits and a concerning rate of suicidal ideation among self-identifying incels (Speckhard et al., 2021). Thus, in addition to adhering to a shared ideology, incels often have common grievances and similar experiences with various psychological challenges.

Interestingly, despite the recurring themes of loneliness and suffering that circulate within incel forums, many users report experiencing some positive effects from their participation. In a study of 250 self-identified incels, participants noted that engaging with others in these forums helped them feel less lonely and fostered a sense of belonging and being understood (Speckhard et al., 2021). This suggests that, driven by feelings of isolation, individuals may turn to these communities in search of connection and mutual understanding. However, despite these positive effects, research also indicates that socially isolated individuals who engage heavily with online platforms are at a heightened risk of developing problematic internet use habits (Maxwell et al., 2020). These may include spending increasing amounts of time online and experiencing distress when offline. Such growing dependence on online engagement can further reinforce social isolation, resulting in increased self-reported anxiety and depression, and potentially trapping individuals in a cycle of deepening involvement within these communities (Maxwell et al., 2020). Moreover, such prolonged online connectivity fosters a space where incels can reinforce their grievances and problematic beliefs, resulting in a self-reinforcing echo-chamber which serves to keep members engaged in the community (De Roos et al., 2024).

The present study

Although some studies have attempted to understand how self-identified incels disengage from the community (e.g., Burns & Boislard, 2024), the psychological and linguistic factors underpinning the process of disengagement and re-engagement remain largely unexplored. Existing literature largely relies on static, cross-sectional analyses of online extremist rhetoric, providing only a snapshot of user psychology. This prevents practitioners from identifying the dynamic shifts that precede significant behavioural changes, such as leaving the community or escalating risk. The primary novel contribution of this study is the development of a longitudinal, linguistic framework that predicts engagement outcomes based on user trajectory. It is essential that we understand why some incels remain within the communities while others eventually leave, and why some incels who leave end up returning.

Given that the incel community operates almost exclusively online (Speckhard et al., 2021), analysing the language used in forum discussions provides a unique avenue for tracking the psychological variables associated with engagement. Building on prior research that has successfully used linguistic analysis to examine mental health patterns among social media users (e.g., Chandra Guntuku et al., 2019; Lyons et al., 2018; Silveira et al., 2021), the present study used Linguistic Inquiry and Word Count (LIWC) dictionaries to investigate the linguistic expression of psychological variables across different patterns of engagement in an incel forum. Our analyses focused on three main categories:

1. Grievances: dictionaries measuring desperation, frustration hate, jealousy, loneliness, paranoia, suicide, threat, and violence, which reflect the core themes of perceived injustice and marginalisation discussed in the literature.
2. Emotions: dictionaries for negative affect such as sadness, anger and anxiety, which are consistently linked to susceptibility to extremist communities, and positive affect more broadly.
3. Social and cognitive dynamics: dictionaries for Analytic Thinking, Clout, Authenticity, and Tone allow us to explore the cognitive and social processes inherent to these communities, such as how echo chambers may influence critical thought, how individuals gain status, and how the overall emotional valence of language shifts over time.

Specifically, we addressed two main research questions. First, how do users who permanently leave the forum (Desisters) differ from those who remain (Persisters) and those who leave but later return (Returners), in terms of psychological and linguistic markers overall and throughout their engagement? And second, can these markers be identified across key timepoints, namely a user's initial period of forum activity (Baseline), their final period of activity before disengagement (Pre-Gap) and their first period of activity upon re-engagement (Post-Gap)?

We hypothesised that, H1) compared to Desisters, Persisters and Returners would exhibit greater average levels of negative affect and grievances, and (H2) the trajectory of negative affect and grievances would differ significantly among the three groups. Additionally, (H3) we expected users would express lower levels of negative affect and grievances immediately before disengagement (Pre-Gap) than at initial forum activity (Baseline). Finally, (H4) we expected upon re-engagement (Post-Gap) negative affect and grievances would return to levels comparable to those observed at Baseline.

Transparency and openness

This study was not preregistered. Due to the sensitive nature of the online community data and to safeguard the privacy and anonymity of participants, the de-identified data and associated code are not publicly archived. While efforts have been made to de-identify the data, the unique linguistic patterns and specific forum interactions within this context could pose a potential risk of re-identification,

even if minimal. This approach aligns with ethical guidelines for handling data from potentially vulnerable online populations. The code and de-identified data will be made available to qualified researchers upon reasonable request to the corresponding author, provided they agree to adhere to strict ethical guidelines, data security protocols, and relevant institutional review board (IRB) approvals. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. This study was approved by the authors' university's Ethics Committee (Reference: ETH2324-0165).

Limitation statement

Consistent with journal requirements, we acknowledge the importance of reporting participant demographic data. However, due to the nature of our dataset, which consists of publicly available online forum posts, detailed demographic information on participants (including racial/ethnic identification, cultural/geographic background, sex/gender, or measures of income, education, and socioeconomic status) was not directly collected and is thus unavailable. This limitation is inherent to analyses relying on large-scale, naturally occurring online behavioural data, where user anonymity and the ethical constraints of data collection preclude gathering such sensitive personal information. While we were able to analyse changes in linguistic patterns over time, the absence of these demographic variables means we cannot assess potential moderating effects of identity or socioeconomic factors on the observed trajectories. Future research leveraging alternative data collection methods may explore these important considerations.

Methods

Data collection and preparation

The data for this study were obtained from the public section of the most active incel forum, www.incels.is. Using Python libraries BeautifulSoup and Requests, we scraped the entire publicly accessible portion of the forum, from its first recorded comment in 2017 to the end of November 2023. For each comment, additional metadata were collected, including the username, user join date, total number of posts by the user (at the time of each comment), and the timestamp of the comment. To ensure privacy of users, usernames were anonymised before the data were analysed. Although the dataset includes all publicly accessible comments on the forum during the study period, it may not fully represent the broader incel community or private interactions, which could limit generalisability. By focusing on the public content, the study ensures transparency and reproducibility, though findings should be interpreted in the context of publicly visible forum activity. This process produced a dataset comprising 5,717,054 comments spanning a six-year period.

Analytical approach

To investigate linguistic indicators of psychological states across patterns of forum engagement, we conducted a series of data preparation and statistical modelling steps using R. Our approach involved defining user activity, classifying users into engagement groups, and conducting longitudinal analyses.

Defining engagement and disengagement

First, to determine a threshold for "active engagement", we analysed weekly posting patterns. Given the highly skewed distribution of weekly posts (median = 23.3, mean = 638.86), we opted for the median as a more robust indicator of typical activity. Consequently, a week of active engagement was defined as contributing at least 25 posts. An 'active period' was subsequently defined as a minimum of ten weeks of active engagement within any consecutive 14-week window. This window was chosen to allow for minor fluctuations in daily posting while still capturing sustained activity.

To define a period of disengagement, we sought a threshold that indicated a significant cessation of activity rather than typical fluctuations. While the mean inter-post interval was 6.26 days, we observed that a three-month (approximately 90-day) period of complete inactivity was a robust indicator of a user's temporary or permanent departure from the forum. This duration was chosen to distinguish between short breaks and genuine disengagement, allowing ample time for users to return if their absence was temporary.

Our operational definitions of engagement and disengagement were informed by prior longitudinal analyses of online extremist communities, which consistently show that participation follows a highly skewed distribution wherein a small subset of highly active users drive most discourse while the majority exhibit rapid attrition after initial involvement (Helm et al., 2024; Kleinberg et al., 2021; Scrivens et al., 2020, 2022). Helm et al. (2024), for instance, identified low-, moderate-, and high-engagement trajectory groups characterised by rapid decline versus sustained, high-frequency posting across 21 months—demonstrating that sustained, high-volume activity is both rare and behaviourally meaningful. Similarly, Scrivens and colleagues (2020, 2022) found that "super-posters" and "committed users" accounted for a disproportionate share of extremist content production. Together, these findings justify the use

of an upper-tail activity threshold (≥25 posts per week) to define active engagement, rather than relying on central-tendency measures.

To identify disengagement, we adopted a ≥90-day inactivity window, consistent with temporal patterns observed in comparable datasets where absences exceeding three months typically signify sustained withdrawal rather than temporary fluctuation (Helm et al., 2024; Kleinberg et al., 2021). This approach ensured that engagement reflected genuine behavioural investment and that disengagement captured meaningful detachment from community discourse. Collectively, our definitions align with established empirical typologies of extremist forum participation, providing both methodological transparency and a replicable framework for future research.

User Group Classification

Based on these definitions, we classified users into three distinct engagement groups for analysis:

1. **Persisters ($n = 713$):** maintained active engagement (as defined above) without any subsequent 3-month gap following their initial active period until the dataset cutoff.
2. **Desisters ($n = 186$):** had a period of active engagement (at least 25 posts per week for at least one stretch of ten out of fourteen weeks) but stopped posting for three months and did not return to the forum before the dataset cutoff (three months before the data were scraped).
3. **Returners ($n = 52$):** stopped posting to the forum for at least three months but returned at some point and upon their return showed active engagement (25 posts per week) for at least eight weeks. Here, we defined an active re-engagement period as eight rather than ten weeks of activity for two reasons, first, our dataset has a fixed end date. A stricter ten-week rule would unfairly exclude users who genuinely re-engaged but had less than ten weeks to post before the data collection concluded. Second, with only 52 Returners, this was our smallest group. A stricter definition would risk a sample size too small for robust statistical analysis. The eight-week threshold provides a larger, more powerful sample, making the longitudinal analysis of this key group more reliable.

These classifications were then merged back into the full dataset and only comments from users classified into one of these three groups of engagement were retained.

Data preparation and analysis

To prepare the data for analysis, all comments with fewer than 10 words were excluded, following guidelines from the Linguistic Inquiry and Word Count (LIWC) tool, which recommends a minimum word count to ensure reliable content analysis (Boyd et al., 2022). Following this, the Persisters had a total of 1,456,183 comments, the Desisters 328,834 comments, and the Returners 224,858 comments.

All LIWC-derived psychological variables were log-transformed using $\log(x+1)$ prior to modelling. This procedure was necessary to normalise the data and ensure it met the underlying assumptions of the linear models. Similarly, continuous predictors (Time, Days on forum, and Number of posts) were standardised to improve model convergence and aid in the interpretation of the results. Given the hierarchical and longitudinal nature of our data, model assumptions were primarily assessed by examining the distribution of the model residuals rather than the raw outcome variables, as is standard practice for linear mixed-effects models (e.g., Gelman & Hill, 2007). This approach accounts for the non-independence of observations and the intrinsic non-normal distribution often seen in language data.

To test our hypotheses, we fitted a series of linear mixed-effects models (LMMs) for each LIWC variable using the lme4 and lmerTest packages in R. We chose LLMs because this method is specifically designed to handle nested, repeated-measures data (multiple comments nested within individual users). This approach is necessary because it accounts for the non-independence of data points from the same individual, providing a more robust and reliable estimate of change over time. The general structure of the models included:

- Fixed effects: the primary variables of interest (LIWC summary variables, emotions and grievances), and covariates (Days on Forum and Number of Posts).
- Random intercepts: Used for each user to account for the unique baseline level of each individual.

Models were fitted using maximum likelihood (ML) estimation. Given the size of the dataset, we relied on the asymptotic approximation method for estimating the degrees of freedom for fixed effects. This method ensures model stability and feasible computation time for large datasets whilst still providing reliable statistical inference.

To test H1 and H2, we analysed the full posting history of all three engagement groups. The LMM for each LIWC variable included Engagement group as a fixed effect, Time as a continuous covariate, and the Engagement x Time

interaction term. We also included Days on forum (the total days a user was active) and Number of posts (the total number of posts) as covariates to control for these aspects of user activity. To test H1 we performed post-hoc pairwise comparisons on the main effect of Engagement group to compare the estimated marginal means of each group. This assesses the overall average level of each linguistic variable across the three groups, adjusting for multiple comparisons with a Tukey adjustment. To test H2, we used the emtrends function to perform post-hoc comparisons with Bonferroni corrections on the slopes of the Time variable for each Engagement group. This directly tested whether the rate of change in language used over time differed across Engagement groups.

To test H3 and H4, we defined three critical temporal segments for all users with a gap in their posting history. These segments were based on a user's absolute first post on the forum, not the start of their identified active period. This approach was chosen to capture the earliest available linguistic baseline for each user. The segments were:

1. **Baseline:** The first 14 days of a user's entire posting history on the forum.
2. **Pre-Gap:** The last 14 days of posting prior to their 3-month posting gap.
3. **Post-Gap:** The first 14 days of posting following their return from the gap.

The resulting dataset for this analysis included posts from Persisters (for the baseline period only), Desisters (for baseline and pre-gap periods), and Returners (for all three periods). We fitted a series of linear mixed-effects models (LMMs) for each LIWC variable, modeling scores as a function of the period (Baseline, Pre-Gap, Post-Gap) and including a random intercept for each user. This repeated measures approach allowed us to assess whether specific psychological markers changed across these critical time points, with the comparisons necessarily being a mix of within-user (for Desisters and Returners) and between-group (e.g., comparing a Persister's baseline to a Desister's baseline). Post-hoc comparisons were performed using Tukey's HSD adjustment, with the Baseline period serving as the reference level.

Dictionary variables

To conduct this research, we utilised several validated dictionaries. We began with the summary variables provided by LIWC for each text analysis. Then, given incels' focus on self-perceived grievances, we specifically selected a subset of grievance-related dictionaries for further examination.

LIWC variables

LIWC includes a range of summary variables (for validation, see Boyd et al., 2022). The first of these is word count, measured here at the level of individual comments. Analytical Thinking captures the degree to which language reflects logical reasoning; lower scores suggest more intuitive and personal expression, while higher scores indicate more structured, analytical thought. Clout reflects perceived social status as conveyed through language. High Authenticity scores are associated with more spontaneous and less self-conscious communication, whereas lower scores imply greater self-monitoring or linguistic filtering. Lastly, Tone represents the emotional valence of language, with higher scores indicating a more positive emotional tone and lower scores suggesting negativity.

In addition to these summary measures, LIWC offers dictionaries that capture various emotional states. For the purpose of this research, we included dictionaries for Anxiety (e.g., worry, fear, afraid, nervous), Anger (e.g., hate, mad, angry, frustr*), and Sadness (e.g., sad, disappoint*, cry).

Grievances

Finally, we incorporated the following sub-dictionaries from the Grievance Dictionary developed by van der Vegt and colleagues (2021): Desperation (e.g., sorrow, last chance, urgent), Frustration (e.g., annoyed, problem, powerless), Hate (e.g., enemy, loathe, hatred), Jealousy (e.g., cheat, resent, bitter), Loneliness (e.g., disconnected, nobody, abandon), Paranoia (e.g., suspicious, conspiracy, suspect), Suicide (e.g., die, overdose, last resort), Threat (e.g., warn, danger, unsafe), and Violence (e.g., bloodshed, fight, bullet).

Results

We conducted two sets of linear mixed-effects models (LMMs) to investigate the linguistic expression of psychological variables across different patterns of forum engagement. The first set of models examined the average levels and long-term trajectories of dictionary variables, while the second set focused on within-user changes during critical time periods.

Between-group differences and longitudinal trajectories

The LMMs investigating overall group differences are displayed in Table 1. They revealed distinct patterns in both the baseline levels and trajectories of dictionary variables across the three Engagement groups (Desisters, Persisters and

Table 1 Linear Mixed Models of Baseline and Trajectory for each Dictionary Variable controlling for number of days and number of posts on forum

Grievances	Desperation		Frustration		Hate		Jealousy		Loneliness		Paranoia		Suicide		Threat		Violence	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Intercept	.270***	.005	.107***	.003	.186***	.004	.102***	.003	.358***	.005	.144***	.003	.287***	.006	.137***	.004	.270***	.006
Persist	-.010*	.005	-.003	.003	.000	.005	-.007*	.003	-.011*	.005	-.004	.003	.001	.006	.004	.004	.008	.006
Return	-.001	.009	-.003	.005	-.006	.008	-.007	.005	-.007	.009	-.001	.006	.003	.011	<.001	.007	.012	.011
Time	.002	.002	.001	.001	.010***	.002	.002	.001	.006**	.002	.002	.001	.003	.002	.010***	.001	.008***	.002
Days on forum	.004	.003	.004*	.002	.002	.002	.003*	.001	.006*	.003	.003	.002	.006	.003	<.001	.002	.003	.003
Posts on forum	.003	.004	.001	.002	-.003	.003	.002	.002	.007	.004	.002	.002	-.006	.005	-.002	.003	-.003	.004
Persist x time	-.010***	.002	-.006***	.001	-.011***	.002	-.007***	.001	-.017***	.002	-.007***	.001	-.013***	.002	-.008***	.001	-.011***	.002
Return x time	-.008***	.002	-.001	.001	-.003	.002	-.001	.001	-.011***	.002	<.001	.002	-.002	.002	-.004*	.002	-.002	.002
Summary / Affect	Word Count		Positive		Anxiety		Anger		Sadness		Analytic		Clout		Authentic		Tone	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Intercept	3.346***	.023	.255***	.005	.038***	.001	.100***	.003	.048***	.002	2.935***	.025	3.137***	.027	3.128***	.021	2.757***	.013
Persist	-.049*	.024	-.008	.006	−0.001	.001	.001	.003	-.003	.002	.030	.026	-.004	.028	-.039	.022	-.024	.014
Return	-.023	.043	-.003	.010	−0.001	.003	.001	.006	-.004	.004	.025	.046	-.024	.050	-.048	.040	-.005	.025
Time	.047***	.002	-.007***	.002	0.000	.001	.002	.001	.001	.001	.028***	.004	.007	.004	.018***	.004	-.020***	.004
Days on forum	.003	.012	.004	.003	<.001	.001	-.001	.002	.003**	.001	.016	.013	.013	.014	-.001	.011	.004	.007
Posts on forum	.078***	.018	-.004	.004	<.001	.001	.003	.002	-.001	.001	-.025	.019	.038	.020	.018	.016	-.006	.010
Persist x time	-.060***	.002	-.004	.002	-.001	.001	-.002	.001	-.004	.002	-.010*	.004	.012*	.005	-.042***	.005	.010*	.004
Return x time	-.046***	.002	-.002	.002	.001	.001	-.002	.001	-.002	.002	-.031***	.004	.013*	.005	-.027***	.005	-.005	.005

Note. Dependent variables are log-transformed (log(x + 1)). Continuous predictors (Time, Days on forum, Posts on forum) are z-scored. Coefficients (B) for continuous predictors represent the change in the log-transformed dependent variable for a one-standard-deviation increase in the predictor. While individual coefficients may appear small, particularly for Time and its interactions, they represent cumulative effects over the study period or across the range of the predictor; ***p<.001, **p<.01, *p<.05

Returners), after controlling for Days on forum and Number of posts (H1). For Persisters, baseline levels of desperation, jealousy, and loneliness were lower than for Desisters, and their posts were shorter than those of Desisters. There were no significant baseline differences between Returners and Desisters. The control variable Days on the forum was positively associated with frustration, jealousy, loneliness and sadness, suggesting these emotions were more prevalent the longer users were active on the forum. The number of posts was associated only with increased word count.

The interaction between Engagement group and Time (Table 1) revealed linguistic trajectories differed across groups (H2). Desisters (reference group) showed significant increases over time in hate, loneliness, threat, violence, authenticity and analytical language. Their posts also grew longer over time, whereas their overall tone become more negative. Interestingly, compared to Desisters, Persisters showed various decreases or slower increases over time, in desperation, frustration, hate, jealousy, loneliness, paranoia, suicide, threat, and violence linguistic markers. Over time, compared to Desisters, their posts became shorter, less analytic and authentic, but more indicative of clout. Compared to Desisters, the Returners decreased or increased more slowly over time on desperation, loneliness and threat. Their posts also became shorter, and showed less analytic thought and authenticity but more clout compared to Desisters. Post-hoc comparisons (Tukey-adjusted) revealed some differences in trajectory over time between Persisters and Returners. Returners' frustration, hate, jealousy and loneliness increased or decreased more slowly over time compared to Persisters. Similarly, Returners displayed increased or slower decreases in paranoia, suicide, threat and violent language, their posts became longer and they were less authentic over time compared to Persisters. Their anxiety and sadness increased compared to Persisters, and they became more analytic in their posts over time, whereas their anger decreased. Table 2

Changes at critical time periods

A second set of linear mixed models was used to investigate changes in dictionary variables during specific, critical time windows: the first two weeks of activity (baseline), the last two weeks before a > 3-month gap (pre-gap) and the first

Table 2 Post-Hoc Comparisons of Slopes for each Dictionary Variable

Variable	Contrast	Est	SE	Variable	Est	SE
Desperation	desister—persister	.010***	.002	Word count	.060***	.002
	desister—returner	.008***	.002		.046***	.002
	persister—returner	-.003	.001		-.015***	.002
Frustration	desister—persister	.006***	.001	Analytic	.010	.004
	desister—returner	.001	.001		.031***	.004
	persister—returner	-.006***	.001		.021***	.003
Hate	desister—persister	.011***	.002	Clout	-.012*	.005
	desister—returner	.003	.002		-.013*	.005
	persister—returner	-.008***	.001		-.001	.003
Jealousy	desister—persister	.007***	.001	Authentic	.042***	.005
	desister—returner	.001	.001		.027***	.005
	persister—returner	-.006***	.001		-.015***	.003
Loneliness	desister—persister	.017***	.002	Tone	-.010	.004
	desister—returner	.011***	.002		.031***	.004
	persister—returner	-.005**	.002		.021***	.003
Paranoia	desister—persister	.007***	.001	Positive emotion	-.012*	.005
	desister—returner	.000	.002		-.013*	.005
	persister—returner	-.007***	.001		-.001	.003
Suicide	desister—persister	.013***	.002	Anxiety	.042***	.005
	desister—returner	.002	.002		.027***	.005
	persister—returner	-.010***	.001		-.015***	.003
Threat	desister—persister	.008***	.001	Anger	-.010	.004
	desister—returner	.004	.002		.005	.005
	persister—returner	-.004***	.001		.015***	.003
Violence	desister—persister	.011***	.002	Sadness	.004***	.002
	desister—returner	.002	.002		.002	.001
	persister—returner	-.01***	.00		-.003***	.001

Note. Dependent variables are log-transformed (log(x + 1)). The Time variable (time_days_scaled) is z-scored. Estimates (Est.) represent the difference in the slope of the log-transformed dictionary variable per one standard-deviation increase in Time between the contrasted engagement groups; p-values were derived using an asymptotic approximation which uses a z-ratio; ***$p < .001$, **$p < .01$, *$p < .05$

two weeks after returning (post-gap) (H3, H4). Results of the linear mixed models are displayed in Table 3. Post hoc-comparisons (Table 4) revealed three significant changes across the critical time periods, all pertaining to the comparison between a user's initial activity and their post-disengagement period. Both desperation and frustration scores were significantly lower in the post-gap period compared to baseline. Specifically, desperation decreased from baseline to post-gap ($Est.=0.06$, $SE=0.03$, $p<0.05$), and frustration followed a similar pattern ($Est.=0.04$, $SE=0.02$, $p<0.05$). Similarly, Analytic scores also showed a significant increase, being higher in the post-gap period compared to baseline ($Est.=-0.16$, $SE=0.06$, $p<0.05$). Scores for hate in the post-gap period were higher than in the pre-gap period, but this difference did not reach significance. No other dictionary variables showed significant changes across these critical time points.

Discussion

This study aimed to investigate linguistic markers of psychological states among incel forum users, to understand differences across distinct engagement groups and their psychological trajectories over time. Our findings suggest that the psychological profiles of incels are not homogeneous, but rather reflect significant distinctions based on their patterns of engagement within the community.

We broadly hypothesised there would be overall differences between groups (H1), and we found partial support for this. Persisters began the study with significantly lower levels of desperation jealousy and loneliness compared to Desisters. These baseline levels suggest a milder psychological vulnerability in the group that eventually persists in their engagement. On the contrary, our results provided strong support for our second hypothesis (H2) that psychological trajectories differ significantly between groups. Desisters' language became increasingly grievance-focused and hostile over time, whereas Persisters' language showed more stable or even positive shifts. Finally, our analysis of critical time periods provided mixed support for our hypotheses (H3 and H4). We found that a break from the forum might temporarily reduce feelings of desperation and frustration as these were lower at post-gap than baseline level, but post-gap comments displayed lower levels of analytic reasoning. Upon re-engagement, none of the other variables showed significant changes compared to baseline or pre-gap levels (H4).

The observation that Desisters' language became increasingly negative over time, as marked by significant increases in hate, loneliness, threat, and violence-related language, alongside increases in analytic reasoning and authenticity.

Table 3 Linear Mixed Models of Critical Time Periods for each Dictionary Variable controlling for number of days and number of posts on forum

Grievances	Desperation		Frustration		Hate		Jealousy		Loneliness		Paranoia		Suicide		Threat		Violence	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Intercept	.277***	.003	.117***	.002	.191***	.003	.107***	.002	.365***	.004	.151***	.003	.312***	.004	.138	.003	.280	.004
Pre-Gap	-.037	.023	-.014	.015	.014	.020	.001	.015	-.027	.026	-.007	.017	-.010	.025	.000	.017	.011	.024
Post-Gap	-.063*	.026	-.043*	.018	-.029	.023	-.015	.017	-.033	.029	-.018	.020	-.034	.029	-.004	.020	.012	.028
Days on forum	.015*	.006	.011**	.004	.009	.005	.004	.004	.011	.007	.004	.004	.010	.006	.004	.004	-.002	.006
Posts on forum	.007*	.003	.002	.002	<.001	.003	.001	.002	.006	.004	.001	.002	.000	.004	-.001	.002	.001	.004
Summary /	Word Count		Positive		Anxiety		Anger		Sadness		Analytic		Clout		Authentic		Tone	
Affect	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Intercept	3.360***	.011	.276***	.004	.042***	.001	.103***	.002	.052***	.002	2.947***	.013	3.075***	.014	3.164***	.012	2.739***	.009
Pre-Gap	.009	.027	-.022	.024	-.009	.009	-.020	.015	-.004	.010	.075	.051	.067	.059	.013	.057	.025	.053
Post-Gap	-.038	.034	-.040	.027	.001	.010	.003	.017	-.014	.012	.158	.062	.046	.071	-.103	.068	.093	.063
Days on forum	.028***	.007	.010	.006	.000	.002	-.001	.004	.007*	.003	-.014*	.013	-.001	.016	.022	.015	-.023	.014
Posts on forum	-.007	.006	-.002	.004	.001	.001	.002	.002	-.001	.001	.005	.010	-.003	.012	.015	.010	.004	.008

Note. Dependent variables are log-transformed ($\log(x+1)$). Days on forum and Posts on forum are z-scored. Categorical time points (Pre-Gap, Post-Gap) are compared to the reference category (Baseline). Coefficients (B) represent the change in the log-transformed dependent variable for a one-standard-deviation increase in Days on forum or Posts on forum, or the difference in the log-transformed dependent variable between the specified time point and the reference time point. ***$p<.001$, **$p<.01$, *$p<.05$

Table 4 Post-Hoc Comparisons of Critical Time Periods for all Dictionary Variables

Variable	Contrast	Est	SE	Variable	Est	SE
Desperation	Baseline – Pre-Gap	.037	.023	Word count	-.009	.027
	Baseline – Post-Gap	.063*	.026		.038	.034
	Pre-Gap – Post-Gap	.025	.023		.047	.026
Frustration	Baseline – Pre-Gap	.014	.015	Analytic	-.075	.051
	Baseline – Post-Gap	.043*	.018		-.158*	.062
	Pre-Gap – Post-Gap	.029	.015		-.083	.049
Hate	Baseline – Pre-Gap	-.014	.020	Clout	-.067	.059
	Baseline – Post-Gap	.029	.023		-.046	.071
	Pre-Gap – Post-Gap	.043	.019		.021	.057
Jealousy	Baseline – Pre-Gap	-.001	.015	Authentic	-.013	.057
	Baseline – Post-Gap	.015	.017		.103	.068
	Pre-Gap – Post-Gap	.017	.014		.116	.055
Loneliness	Baseline – Pre-Gap	.027	.026	Tone	-.025	.053
	Baseline – Post-Gap	.033	.029		-.093	.063
	Pre-Gap – Post-Gap	.006	.025		-.068	.052
Paranoia	Baseline – Pre-Gap	.007	.017	Positive emotion	.022	.024
	Baseline – Post-Gap	.018	.020		.040	.027
	Pre-Gap – Post-Gap	011	.017		.018	.023
Suicide	Baseline – Pre-Gap	.010	.025	Anxiety	.009	.009
	Baseline – Post-Gap	.034	.029		-.001	.010
	Pre-Gap – Post-Gap	.023	.024		-.009	.009
Threat	Baseline – Pre-Gap	.000	.017	Anger	.020	.015
	Baseline – Post-Gap	.004	.020		-.003	.017
	Pre-Gap – Post-Gap	.004	.017		-.023	.015
Violence	Baseline – Pre-Gap	-.011	.024	Sadness	.004	.010
	Baseline – Post-Gap	-.012	.028		.014	.012
	Pre-Gap – Post-Gap	-.001	.023		.010	.010

Note. Dependent variables are log-transformed (log(x + 1)). Estimates (Est.) represent the difference in the mean of the log-transformed dictionary variable between the contrasted time periods. p-values were derived using an asymptotic approximation which uses a z-ratio; ***$p < .001$, **$p < .01$, *$p < .05$

Their posts also grew longer over time, whereas their overall tone became more negative. This concurrent escalation in negative markers and linguistic complexity may reflect a process of disillusionment and movement burnout. This aligns with existing literature, which suggests that disillusionment occurs when members' expectations of community support or shared purpose clash with the community's actual norms and behaviours (Simi et al., 2019). Over time, this discrepancy can contribute to psychological strain and burnout (Gaudette et al., 2022). Crucially, this implies that disengagement from the forum does not necessarily indicate emotional recovery. Rather, it may signal that the psychological cost of continued participation has become unsustainable, compelling members to withdraw despite persisting, or even intensifying, negative emotions. Additionally, Desisters' posts grew longer over time. This concurrent increase in word count could imply that as emotional burden intensifies, users engage in more elaborative self-expression, potentially as a form of emotional offloading. This is consistent with findings from expressive writing research, where longer narratives correlate with the processing of distress (Tausczik & Pennebaker, 2010). Thus, the rise in wordcount suggests a final intensification of communicative effort before disengagement.

Persisters exhibited a more stable, or at least less negative, emotional trajectory than Desisters, suggesting their engagement may serve as a relatively effective coping mechanism. Compared to Desisters, Persisters showed various decreases or slower increases over time in desperation, frustration, hate, jealousy, loneliness, paranoia, suicide, threat, and violence linguistic markers. This aligns with existing literature indicating that some incels benefit from positive aspects of community involvement, such as fostering a sense of belonging and feeling understood (Speckhard et al., 2021). Additionally, Persisters showed an increase in Clout over time, without significant corresponding changes in Authenticity or Tone. Further, we observed a simultaneous significant decrease (or slower increase) in Analytic thought and Authenticity compared to Desisters. This suggests growing social status or authority within the forum, a pattern well-documented in other extremist online communities where a core of highly active, tightly connected members disproportionately influence conversation and information flow; a pattern that has been documented in

other extremist communities (Asatani et al., 2021). Research on online social dynamics also indicates that status mechanisms, such as Reddit's comment voting systems, critically motivate user participation by reinforcing acceptance within the community, thereby influencing radicalisation trajectories (Habib et al., 2022). Moreover, studies of extremist and fringe forums find that a small subset of highly active or veteran users often gain informal authority, acting as a social reward that sustains participation (Baele et al., 2022; Rothut et al., 2024). Together, these findings suggest that Persisters' rising Clout reflects social reinforcement for loyalty, helping to maintain their engagement, even as emotional indicators remain stagnant or worsen and their critical thinking reflected in language through their analytic and authenticity levels decreases over time.

The Returners in our sample appeared to fall somewhere between Desisters and Persisters; they are not as distressed as the Desisters who end up leaving the forum, but they also do not seem to gain the social benefits of engagement in the same way Persisters do. Where Desisters experience a steep emotional devolution consistent with burnout, the Returners show a flatter decline, which may explain why their disengagement is not permanent. Their positive Clout trajectory compared to Desisters suggests they are gaining some social status within the community. As such, they are uniquely vulnerable to a paradoxical cycle of disengagement due to the emotional burden and re-engagement due to the community's lingering appeal. The break provides temporary relief, but it does not resolve the underlying issues that led them to the forum in the first place.

We found that a break from the forum did lead to significant short-term psychological shifts, specifically a significant decrease in desperation and frustration scores in the post-gap period compared to baseline. However, when users returned to the forum after a break, they showed a significant increase in Analytic thinking compared to their baseline. The return thus marks a re-adoption of the ideology in a more entrenched and articulated way, that is consistent with the effects seen in group polarisation. This increase in structured thought upon re-engagement is consistent with the effects seen in ideological entrenchment, suggesting the Returners have processed and integrated the core grievances into a more linguistically complex framework. During their disengagement, Returners' core beliefs were likely challenged and condemned by the general public, creating a state of cognitive dissonance (Festinger, 1957). Upon their return, this dissonance is lifted by re-engaging with the ideology uncritically, a process that represents a coping mechanism of eliminating internal conflict by embracing the group ideology and reaffirming their identity within the group.

A similar process can be observed for Hate, which was also more pronounced upon re-engagement than at baseline level or leading up to disengagement, though this difference was not significant. This suggests that while disengaged, the beliefs lingered and did not diminish. To the contrary, the user may be doubling down to fit back into the fold, or it may be that in their absence, they received pushback on their beliefs, which deepened them. This is consistent with Social Identity Theory and the concept of group threat (Tajfel & Turner, 1979), where criticism from an out-group strengthens commitment to the in-group and increases animosity toward the out-group. The Returners thus show a unique trajectory; the emotional burden of engagement leads them to disengage, but without a fundamental shift in underlying grievances or cognitive patterns, the individual remains vulnerable to re-radicalisation, or a reversal of the disengagement process.

A layered view of radicalisation

This integrated framework is empirically grounded in the distinct longitudinal patterns observed: the sustained engagement of Persisters is supported by a confluence of social rewards (rising Clout) and cognitive shifts (a relative decrease in Analytic thought over time), while the disengagement of Desisters is preceded by a critical escalation in grievance-focused emotional markers (including hate, loneliness, and threat) that suggests psychological and communicative burnout. Beyond the isolated trends in emotions or cognitive markers identified by our findings, how these factors interact to drive sustained engagement and radicalisation is of interest. Our results point to a layered model of engagement where core affective states, dynamic social grievances and cognitive processes interact to create a self-perpetuating cycle (Fig. 1), consistent with multi-factor models of radicalisation (Borum, 2011; McCauley & Moskalenko, 2008).

Looking at emotions, the incel forum users in our study show certain negative emotions such as anger and anxiety that appear to be stable and trait-like, whereas others such as sadness, loneliness and jealousy appear more dynamic and fluctuate with engagement. Such a distinction mirrors one made in clinical psychology between chronic core affectivity and situation-specific emotional responses (Zelenski & Larsen, 2000). The stable anger and anxiety may represent a chronic state of emotional dysregulation, which form a pre-existing vulnerability that makes individuals susceptible to extremist ideologies (Borum, 2014). The online community then acts as a powerful amplifier, giving a clear focus to this underlying distress by framing it as externally-driven grievances (e.g., loneliness is "caused" by women, jealousy is "justified" by social unfairness). The community's validation of these grievances is particularly appealing to

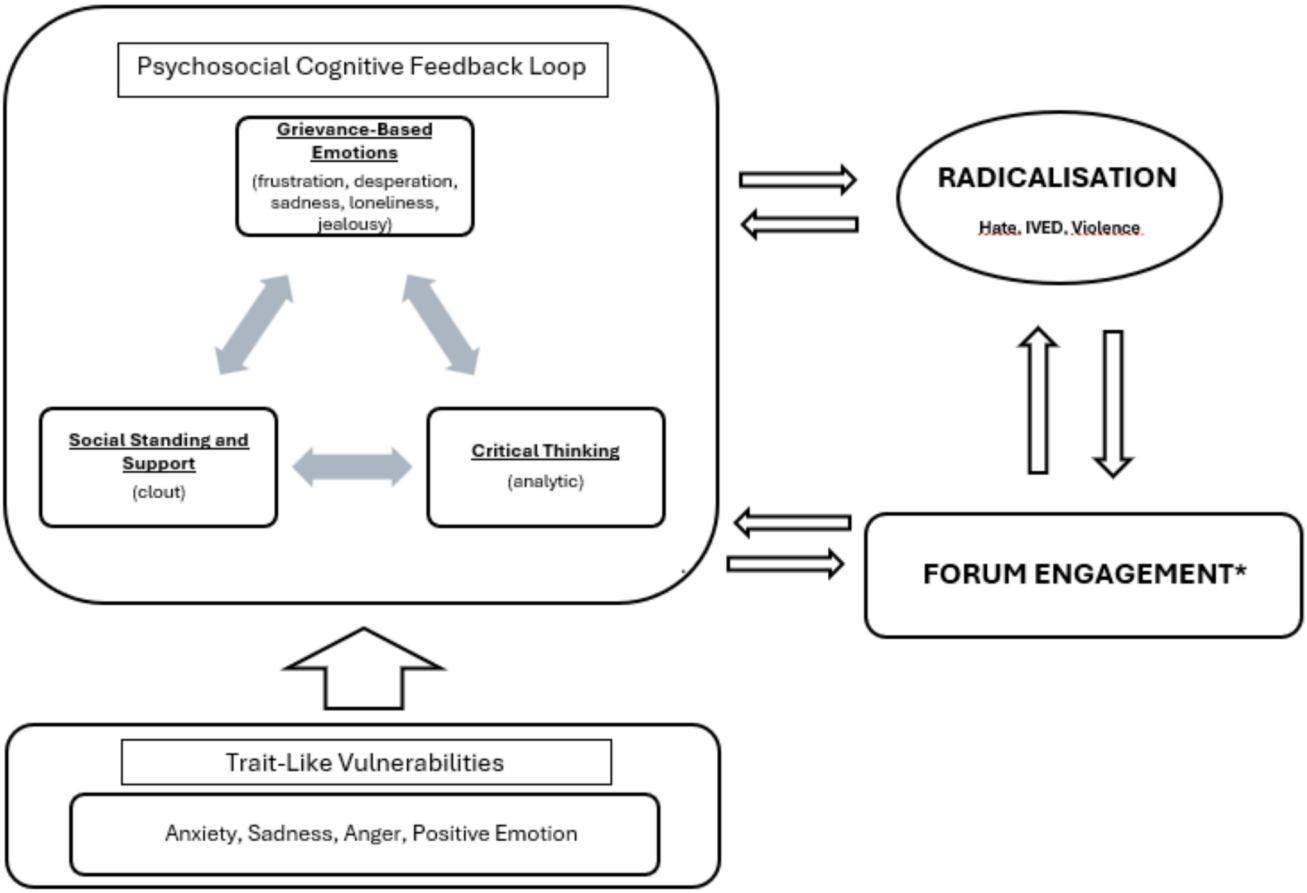

Fig. 1 Proposed Layered Model of Engagement and Radicalisation

individuals who experience chronic distress and perceived marginalisation (Kruglanski et al., 2014).

This emotional vulnerability interacts with cognitive changes. Although chronic emotional distress does not constitute a diagnosis as such, clinically relevant parallels exist with cognitive deficits. Research shows that sustained negative emotional states and rumination are linked with reduced executive function and cognitive rigidity (Snyder, 2013), manifesting as tunnel vision and a reduced capacity for analytical thinking. Our finding that a decline in analytic thinking is part of a broader pattern including trajectories for hate and clout, supports this interpretation. The decline in critical thinking then indicates a cognitive mechanism that allows the group's ideological hate to take root. This is further supported by literature showing that a higher need for cognitive closure and greater dogmatism are predictive of extremist belief systems (Obaidi et al., 2023). In this model, chronic emotional vulnerability predisposes the individual to seek an ideological escape, which the community provides. Through validation of their grievances and social rewards like clout and a sense of belonging, the community and its ideology becomes more central to the individual. This process is then reinforced by a decline in the very cognitive skills needed to question the ideology, which creates a feedback loop where emotional distress drives engagement, and engagement in turn erodes the capacity for critical thought.

By integrating these findings, we propose a layered model for sustained engagement in online extremist communities. This model suggests that engagement is driven by an interaction of:

1. Emotional appeal: The community offers a sense of belonging and validates the individual's core negative affect and grievances.
2. Social rewards: Members are provided with social reinforcement and status, which makes continued engagement a rewarding experience.
3. Cognitive facilitation: The community fosters an environment where analytic thinking declines, thus reducing the psychological friction of adopting and sustaining an extreme ideology.

This integrated framework offers a more nuanced understanding of online radicalisation than a simple, linear path. It highlights how emotional vulnerability is exploited and reinforced through a combination of social and cognitive

mechanisms, creating a powerful, self-perpetuating cycle of engagement that is difficult to break. This model could provide a useful basis for future research into sustained engagement and re-radicalisation, and offers a valuable starting point for developing more targeted clinical interventions.

Limitations & future directions

Due to the correlational nature of the data, we cannot establish a causal relationship between forum engagement and psychological changes. Although we can identify temporal trajectories, we cannot definitively state that forum engagement causes the observed shifts in cognitive and emotional patterns. They may be a reciprocal or co-occurring process. Whilst the use of a single forum provides a deep, longitudinal dataset, it also means our findings may be specific to the culture and moderation of this particular community and may not generalise to other online communities. Future research should examine multiple platforms, and ideologies to determine whether the model we propose generalises to ideologies that may be less grievance-focused. In a similar vein, next steps should include prospectively validating whether the linguistic trajectories identified here can reliably anticipate changes in psychological engagement over time. Developing and testing risk-scoring algorithms or automated classifiers would allow for systematic evaluation of accuracy and false-positive rates, ensuring any applied use of these markers remains empirically grounded and ethically sound. A critical limitation is the inability to fully account for complex user behaviours associated with account management on the forum. Our analysis is based on unique usernames, but we cannot track a single individual across multiple accounts. This means that a user we have classified as Desister may in fact be a "ghost returner" who re-engages with the community under a new identity. Our data also includes accounts labelled as "Deleted Member", which we cannot definitively distinguish from a banned user. As a result, our findings on the Desister and Returner groups are likely conservative and may not represent the full spectrum of user activity. The inability to link new accounts to old ones means our analysis systematically misses a portion of the returning population, potentially skewing the trajectories we have observed. Finally, it is important to note that our linguistic analyses infer psychological or ideological states indirectly through dictionary-based markers. Although these tools are widely used and validated, they do not provide direct measures of users' internal experiences. Future work could address this by integrating AI-driven natural language processing to detect latent patterns, hidden returners, or emergent psychological signals across platforms. Machine-learning models, such as those used by HerEthical.AI to classify victim-blaming narratives (herEthical.AI, 2025) could be adapted to refine our framework, enabling more dynamic and scalable identification of radicalisation-related discourse and improving both research and prevention applications. Furthermore, we did not control for situational or contextual factors, such as specific forum threads or external events, which may have influenced language use. Therefore, the results should be interpreted with caution, recognising that observed linguistic patterns may reflect both individual tendencies and contextual influences.

Practical implications

The longitudinal and differential findings presented in this study offer a novel psychology-grounded model that can be directly leveraged by law enforcement, counter-terrorism initiatives and mental health practitioners to enhance prevention and intervention efforts. The primary novelty of our work is the demonstration that sustained engagement in online extremism is not a linear process of escalation, but is instead driven by distinct, measurable psychological trajectories. This layered framework therefore enables a more targeted approach to prevention and relapse reduction by mapping the dynamic emotional, social, and cognitive mechanisms that sustain or disrupt engagement.

First, the prevention of ideological entrenchment evident in Persisters. The linguistic trajectory of Persisters, which is marked by rising Clout and a concurrent decline in Analytic thinking, signals a transition from emotional vulnerability to cognitive rigidity reinforced through social validation. For practitioners, this combination indicates successful integration into the extremist social structure, where critical reflection has been supplanted by status-based reward. Intervention efforts should therefore move beyond conventional de-radicalisation models and focus on disrupting the social identity affirmation that sustains engagement. Law enforcement and community-based programmes can target this process by promoting constructive, non-ideological pathways to social recognition, such as mentorship or community status mechanisms that fulfil the same psychosocial needs without reinforcing extremist norms. These strategies collapse the feedback loop of social reward and ideological conformity, increasing the likelihood of disengagement.

Second, the opportunity to capitalise on burnout seen in Desisters as an intervention window. Among Desisters, the pre-disengagement escalation of grievance-based language marked by sharp increases in hate, loneliness, and threat-based language represents a form of movement burnout and emotional overload. This pattern identifies a critical, time-sensitive window for intervention, in which the social and emotional costs of continued participation outweigh its perceived benefits. Monitoring these linguistic shifts can

inform threat-assessment protocols and supportive outreach before individuals fully withdraw. In practice, this could guide both online moderation systems and specialised units to initiate empathetic engagement or referral to mental-health services, capitalising on the temporary openness to alternative narratives that burnout creates.

Finally, preventing relapse and re-radicalisation. For Returners, the re-engagement pattern, characterised by heightened analytic thinking and stable or increased hostility, suggests cognitive consolidation of extremist ideology following exposure to external criticism. This mirrors relapse dynamics in other behavioural domains, where returning individuals often exhibit renewed ideological commitment. Prevention efforts should therefore focus on post-disengagement support that addresses residual grievances and identity needs, rather than treating disengagement as recovery. Practically, multi-agency coordination between law enforcement, social workers, and online-safety practitioners could help maintain contact with individuals after disengagement, offering structured emotional support and community reintegration opportunities to reduce the pull of extremist spaces.

Together, these applications demonstrate that linguistic and psychological markers can serve as dynamic indicators for both escalation and disengagement, providing a valuable tool for early detection, risk assessment, and relapse prevention. By linking emotional distress, social reinforcement, and cognitive change, this study advances a model that is both theoretically novel and operationally relevant. Recognising and responding to these interacting layers can inform data-driven, psychology-informed prevention strategies across policing, clinical, and policy domains, ultimately reducing the persistence and recurrence of extremist engagement online.

Conclusion

The findings of this study provide a novel, clinically relevant framework for understanding the cognitive and emotional layers of online extremism. The identification of distinct psychological trajectories among Persisters, Desisters and Returners suggests that the language they use can serve as a tool for risk assessment. Although this study focused on an incel forum, the proposed layered model could potentially be applied to monitor shifts in engagement or emotional and cognitive patterns across diverse online ideological communities, informing preventive or early intervention strategies across diverse contexts.

Acknowledgements We extend our sincere thanks to Ema Bohmova for her valuable assistance in drafting portions of the manuscript and for her diligent work on background research that contributed to the literature review. Her support was greatly appreciated.

Author contributions Conceptualization: Melissa S. de Roos; Methodology: Melissa S. de Roos; Formal analysis and investigation: Melissa S. de Roos, Giorgia Caon; Writing—original draft preparation: Giorgia Caon; Writing—review and editing: Melissa S. de Roos;

Funding This study did not receive any funding.

Data availability The data that support the findings of this study are available from the authors upon reasonable request.

Declarations

Ethical approval This study was approved by the authors' university's Ethics Committee (Reference: ETH2324-0165).

Consent to participate Publicly available internet data was used in this study, thus informed consent was not sought from individuals.

Consent to publish All listed authors consent to publication.

Conflict of interest The authors declare no conflicts of interest.

References

Andersen, J. C. (2024). The incel subculture: Drifting in and out. *Young*. https://doi.org/10.1177/11033088241296626

Asatani, K., Yamano, H., Sakaki, T., & Sakata, I. (2021). Dense and influential core promotion of daily viral information spread in political echo chambers. *Scientific Reports*. https://doi.org/10.1038/s41598-021-86750-w

Baele, S. J., Brace, L., Coan, T. G., & Naserian, E. (2022). Super- (and hyper-) posters on extremist forums. *Journal of Policing, Intelligence and Counter Terrorism, 18*(3), 243–281. https://doi.org/10.1080/18335330.2022.2103386

Borum, R. (2011). Radicalization into violent extremism i: A review of social science theories. *Journal of Strategic Security, 4*(4), 7–36. https://doi.org/10.5038/1944-0472.4.4.1

Borum, R. (2014). Psychological vulnerabilities and propensities for involvement in violent extremism. *Behavioral Sciences & the Law, 32*(3), 286–305. https://doi.org/10.1002/bsl.2110

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. University of Texas at Austin. https://www.liwc.app

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America, 114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Broyd, J., Boniface, L., Parsons, D., Murphy, D., & Hafferty, J. D. (2022). Incels, violence and mental disorder: A narrative review with recommendations for best practice in risk assessment and clinical intervention. *Bjpsych Advances, 29*(4), 254–264. https://doi.org/10.1192/bja.2022.15

Brugnoli, E., Cinelli, M., Quattrociocchi, W., & Scala, A. (2019). Recursive patterns in online echo chambers. *Scientific Reports*. https://doi.org/10.1038/s41598-019-56191-7

Burns, L.-M., & Boislard, M.-A. (2024). I'm better than this": A qualitative analysis of the turning points leading to exiting inceldom. *Journal of Sex Research*. https://doi.org/10.1080/00224499.2024.2340110. Advance online publication.

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., & López, H. M. H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior, 58*, Article 101608. https://doi.org/10.1016/j.avb.2021.101608

Chandra Guntuku, S., Buffone, A., Jaidka, K., Eichstaedt, J. C., & Ungar, L. H. (2019). Understanding and measuring psychological stress using social media. *Proceedings of the International AAAI Conference on Web and Social Media, 13*, 214–225. https://doi.org/10.1609/icwsm.v13i01.3223

Chua, Y. T. (2025). We want you!" Applying social network analysis to online extremist communities. *Terrorism and Political Violence, 37*(3), 409–423. https://doi.org/10.1080/09546553.2024.2304800

Costello, M., Barrett-Fox, R., Bernatzky, C., Hawdon, J., & Mendes, K. (2018). Predictors of viewing online extremism among America's youth. *Youth & Society, 52*(5), 710–727. https://doi.org/10.1177/0044118x18768115

Costello, W., Whittaker, J., & Thomas, A. G. (2025). The dual pathways hypothesis of incel harm: A model of harmful attitudes and beliefs among involuntary celibates. *Archives of Sexual Behavior, 54*(5), 1815–1836. https://doi.org/10.1007/s10508-025-03161-y

Daly, S. E., & Reed, S. M. (2022). I think most of society hates us": A qualitative thematic analysis of interviews with incels. *Sex Roles, 86*(1–2), 14–33. https://doi.org/10.1007/s11199-021-01250-5

De Roos, M. S., Veldhuizen-Ochodničanová, L., & Hanna, A. (2024). The angry echo chamber: A study of extremist and emotional language changes in incel communities over time. *Journal of Interpersonal Violence, 39*(21–22), 4573–4597. https://doi.org/10.1177/08862605241239451

De Vettor, M., Lo Buglio, G., Barsanti, A., Ciocca, G., Gennaro, A., Goksal, R., Lingiardi, V., Giovanardi, G., & Boldrini, T. (2025). Involuntary celibacy (Incel) identity: A thematic analysis of an online community's beliefs and emotional experiences. *Societies, 15*(2), Article 44. https://doi.org/10.3390/soc15020044

Delaney, T. W., Pollet, T. V., & Cook, C. (2024). The mental well-being of involuntary celibates. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2023.112474

Dilkes, J. (2023). Rule 1: Remember the human. A socio-cognitive discourse study of a Reddit forum banned for promoting hate based on identity. *Discourse & Society, 35*(1), 48–65. https://doi.org/10.1177/09579265231190344

Døving, C. A., & Emberland, T. (2021). Bringing the enemy closer to home: 'Conspiracy talk' and the Norwegian far right. *Patterns of Prejudice, 55*(4), 375–390. https://doi.org/10.1080/0031322X.2021.1909933

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fontanesi, L., Marchetti, D., Cosi, G., Limoncin, E., Jannini, E. A., Verrocchio, M. C., & Ciocca, G. (2024). What does it take to make an incel: The role of paranoid thinking, depression, anxiety, and attachment patterns. *Depression and Anxiety*. https://doi.org/10.1155/2024/5512878

Gaudette, T., Scrivens, R., & Venkatesh, V. (2020). The role of the internet in facilitating violent extremism: Insights from former right-wing extremists. *Terrorism and Political Violence, 34*(7), 1339–1356. https://doi.org/10.1080/09546553.2020.1784147

Gaudette, T., Scrivens, R., & Venkatesh, V. (2022). Disengaged but still radical? Pathways out of violent right-wing extremism. *Terrorism and Political Violence, 35*(8), 1775–1800. https://doi.org/10.1080/09546553.2022.2082288

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Habib, H., Srinivasan, P., & Nithyanand, R. (2022). Making a radical misogynist. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1–28. https://doi.org/10.1145/3555551

Hartmann, D., Wang, S. M., Pohlmann, L., & Berendt, B. (2025). A systematic review of echo chamber research: Comparative analysis of conceptualizations, operationalizations, and varying outcomes. *Journal of Computational Social Science, 8*(2), Article 52. https://doi.org/10.1007/s42001-025-00381-z

Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., Rousseau, C., & Sieckelinck, S. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International Journal of Developmental Science, 12*(1–2), 71–88. https://doi.org/10.3233/dev-170233

Helm, B., Holt, T. J., Scrivens, R., Wojciechowski, T. W., & Frank, R. (2024). Examining the online posting behaviors and trajectories of incel forum members. *Crime & Delinquency*. https://doi.org/10.1177/00111287241301009

Jardine, E. (2019). Online content moderation and the Dark Web: Policy responses to radicalizing hate speech and malicious content on the Darknet. *First Monday*. https://doi.org/10.5210/fm.v24i12.10266

King, M., & Taylor, D. M. (2011). The radicalization of homegrown jihadists: A review of theoretical models and social psychological evidence. *Terrorism and Political Violence, 23*(4), 602–622. https://doi.org/10.1080/09546553.2011.587064

Kleinberg, B., Van der Vegt, I., & Gill, P. (2021). The temporal evolution of a far-right forum. *Journal of Computational Social Science, 4*(1), 1–23. https://doi.org/10.1007/s42001-020-00064-x

Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology, 35*(S1), 69–93. https://doi.org/10.1111/pops.12163

Leite-Mendonca, S., & Boislard, M. A. (2024). "Help me please, I need practical advice": A qualitative exploration of social support dynamics among incels on online forums. *The Canadian Journal of Human Sexuality, 33*(3), 467–482. https://doi.org/10.3138/cjhs-2024-0007

Liu, J., Schwarz, A., Risius, M., Hirschheim, R., & Scotter, J. V. (2025). Conceptualizing echo chambers and information cocoons: A literature review and synthesis of current knowledge and future directions. *The Journal of Strategic Information Systems, 34*(2), Article 101904. https://doi.org/10.1016/j.jsis.2025.101904

Lyons, M., Aksayli, N. D., & Brewer, G. (2018). Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior, 87*, 207–211. https://doi.org/10.1016/j.chb.2018.05.035

Madriaza, P., Hassan, G., Brouillette-Alarie, S., Mounchingam, A. N., Durocher-Corfa, L., Borokhovski, E., Pickup, D., & Paillé, S. (2025). Exposure to hate in online and traditional media: A systematic review and meta-analysis of the impact of this exposure on individuals and communities. *Campbell Systematic Reviews*. https://doi.org/10.1002/cl2.70018

Maxwell, D., Robinson, S. R., Williams, J. R., & Keaton, C. (2020). "A short story of a lonely guy": A qualitative thematic analysis of involuntary celibacy using Reddit. *Sexuality & Culture, 24*(6), 1852–1874. https://doi.org/10.1007/s12119-020-09724-6

McCauley, C., & Moskalenko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence, 20*(3), 415–433. https://doi.org/10.1080/09546550802073367

Mølmen, G. N., & Ravndal, J. A. (2021). Mechanisms of online radicalisation: How the internet affects the radicalisation of extreme-right lone actor terrorists. *Behavioral Sciences of Terrorism and Political Aggression, 15*(4), 463–487. https://doi.org/10.1080/19434472.2021.1993302

Moskalenko, S., González, J. F., Kates, N., & Morton, J. (2022a). Incel ideology, radicalization and mental health. *The Journal of Intelligence, Conflict, and Warfare, 4*(3), 1–29. https://doi.org/10.21810/jicw.v4i3.3817

Moskalenko, S., Kates, N., Fernández-Garayzábal González, J., & Bloom, M. (2022). Predictors of radical intentions among incels. *Journal of Online Trust and Safety*, *1*(3). https://doi.org/10.54501/jots.v1i3.57

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology & People, 28*(3), 607–622. https://doi.org/10.1108/itp-09-2014-0198

Obaidi, M., Anjum, G., Bierwiaczonek, K., Dovidio, J. F., Ozer, S., & Kunst, J. R. (2023). Cultural threat perceptions predict violent extremism via need for cognitive closure. *Proceedings of the National Academy of Sciences, 120*(20), Article e2213874120. https://doi.org/10.1073/pnas.2213874120

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

Reeve, Z. (2019). Engaging with online extremist material: Experimental evidence. *Terrorism and Political Violence, 33*(8), 1595–1620. https://doi.org/10.1080/09546553.2019.1634559

Rieger, D., Kümpel, A. S., Wich, M., Kiening, T., & Groh, G. (2021). Assessing the extent and types of hate speech in fringe communities: A case study of Alt-Right communities on 8Chan, 4Chan, and Reddit. *Social Media + Society*. https://doi.org/10.1177/2056305121105290 6

Rothut, S., Schulze, H., Hohner, J., & Rieger, D. (2024). Ambassadors of ideology: A conceptualization and computational investigation of far-right influencers, their networking structures, and communication practices. *New Media & Society, 26*(12), 7120–7147. https://doi.org/10.1177/14614448231164409

De Roy van Zuijdewijn, J., & Bakker, E. (2016). Analysing personal characteristics of lone-actor terrorists: Research findings and recommendations. *Perspectives on Terrorism, 10*(2), 42–49. http://www.jstor.org/stable/26297552

Schils, N., & Verhage, A. (2017). Understanding how and why young people enter radical or violent extremist groups. *International Journal of Conflict and Violence*, *11*(2). https://doi.org/10.4119/UNIBI/ijcv.473

Schmitz, M., Muric, G., Hickey, D., & Burghardt, K. (2024). Do users adopt extremist beliefs from exposure to hate subreddits? *Social Network Analysis and Mining, 14*(1), Article 22. https://doi.org/10.1007/s13278-023-01184-8

Scrivens, R., Davies, G., Gaudette, T., & Frank, R. (2022). Comparing online posting typologies among violent and nonviolent right-wing extremists. *Studies in Conflict & Terrorism, 48*(4), 369–391. https://doi.org/10.1080/1057610x.2022.2099269

Scrivens, R., Wojciechowski, T., & Frank, R. (2020). Examining the developmental pathways of online posting behavior in violent right-wing extremist forums. *CrimRxiv*. https://doi.org/10.21428/cb6ab371.607f92cd

Silveira, B., Silva, H. S., Murai, F., & Da Silva, A. P. (2021). Predicting user emotional tone in mental disorder online communities. *Future Generation Computer Systems, 125*, 641–651. https://doi.org/10.1016/j.future.2021.07.014

Simi, P., Blee, K., DeMichele, M., & Windisch, S. (2017). Addicted to hate: Identity residual among former white supremacists. *American Sociological Review, 82*(6), 1167–1187. https://doi.org/10.1177/0003122417728719

Simi, P., Windisch, S., Harris, D., & Ligon, G. (2019). Anger from within: The role of emotions in disengagement from violent extremism. *Journal of Qualitative Criminal Justice & Criminology*. https://doi.org/10.21428/88de04a1.7dc6a559

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin, 139*(1), 81. https://psycnet.apa.org/doi/https://doi.org/10.1037/a0028727

Speckhard, A., Ellenberg, M., Morton, J., & Ash, A. (2021). Involuntary celibates experiences of and grievance over sexual exclusion and the potential threat of violence among those active in an online incel forum. *Journal of Strategic Security, 14*(2), 89–121. https://doi.org/10.5038/1944-0472.14.2.1910

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.

Tastenhoye, C. A., Dupré, J., Ross, N. E., Bodnar, T. V., & Friedman, S. H. (2023). Incels: An introduction for mental health clinicians. *Journal of Psychiatric Practice, 29*(5), 384–389. https://doi.org/10.1097/pra.0000000000000738

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54. https://doi.org/10.1177/0261927x09351676

Van der Vegt, I., Mozes, M., Kleinberg, B., & Gill, P. (2021). The grievance dictionary: Understanding threatening language use. *Behavior Research Methods, 53*(5), 2105–2119. https://doi.org/10.3758/s13428-021-01536-2

Waldek, L., Ballsun-Stanton, B., & Droogan, J. (2020). After Christchurch: Mapping Online Right-Wing Extremists. Global Network on Extremism & Technology. https://gnet-research.org/2020/12/14/after-christchurch-mapping-online-right-wing-extremists/?utm_source=chatgpt.com

Walther, J. B. (2022). Social media and online hate. *Current Opinion in Psychology, 45*, Article 101298. https://doi.org/10.1016/j.copsyc.2021.12.010

Williams, T. J. V., Tzani, C., Gavin, H., & Ioannou, M. (2023). Policy vs reality: Comparing the policies of social media sites and users' experiences, in the context of exposure to extremist content. *Behavioral Sciences of Terrorism and Political Aggression*. https://doi.org/10.1080/19434472.2023.2195466

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2021). Cognitive and behavioral radicalization: A systematic review of the putative risk and protective factors. *Campbell Systematic Reviews*, *17*(3). https://doi.org/10.1002/cl2.1174

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality, 34*(2), 178–197. https://doi.org/10.1006/jrpe.1999.2275