

How an interaction between psychological, social, and technical factors can produce radicalisation

– and why they need to be understood in combination

Professor Laura G. E. Smith

Department of
Psychology



UNIVERSITY OF
BATH



Bath Institute
for Digital Security
and Behaviour

How non-ideological extreme violence (NIEV) develops, and how this differs from ideologically-driven or impulsive violence

Starting Assumptions:

- I understand NIEV as extreme violence perpetrated by individuals for whom ideology or a group-based sociopolitical grievance is not a primary driver.
- I also assume NIEV involves developing capability and motivation for violence over time. This differs from impulsive violence, which would be a reaction to an immediate circumstance.
- I will refer to the development of motivation for NIEV as *radicalisation*, meaning extremisation of views around violence, in which an individual comes to believe that violence is acceptable.
- I also note that the vast majority of research on radicalisation to date has assumed either an ideological or group-based sociopolitical motivation.

Traditional approaches to radicalisation imply a linear pathway

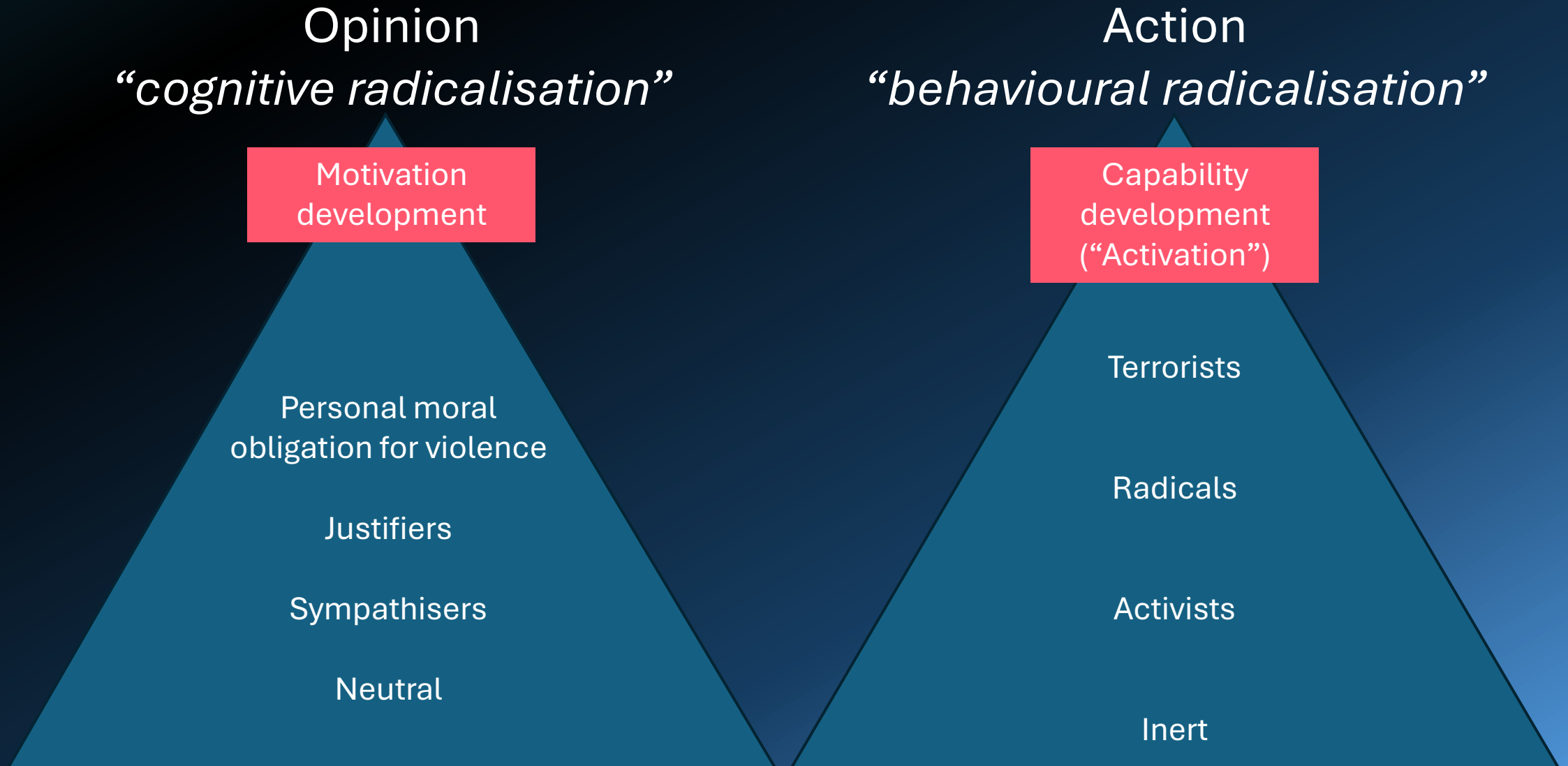
- Staircase is a metaphor, but is misleading and mis-directive:
 - No evidence for steps as distinct stages
 - No evidence for how people move between steps
 - No evidence that people move linearly
 - No evidence of behaviours routinely associated with specific steps
 - Attitude -> behaviour connection is fallacious

- Why we can't predict movement between steps:
 - Low base rate in society
 - Equifinality (many paths to similar outcomes)
- There is no evidence for causal impact of factors with high prevalence in society (like Autism and psychopathology)



"Staircase to terrorism", Moghaddam (2005)

Two-pyramids model (McCauley & Moskalenko, 2017)



Attitudes-behaviours corrective (ABC) model (Khalil et al., 2019)

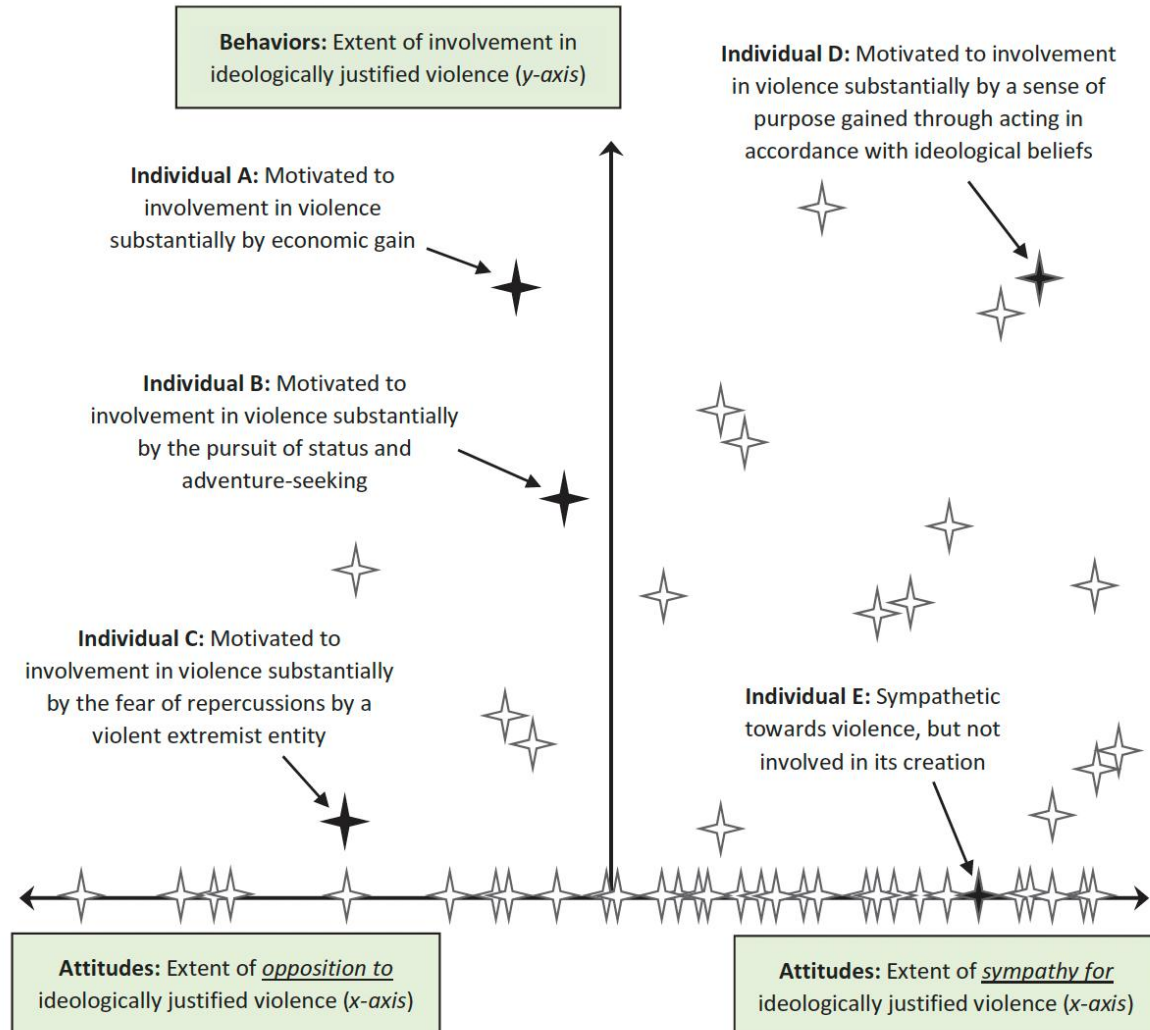


Figure 1. The partial disconnect between attitudes and behaviors.²⁹

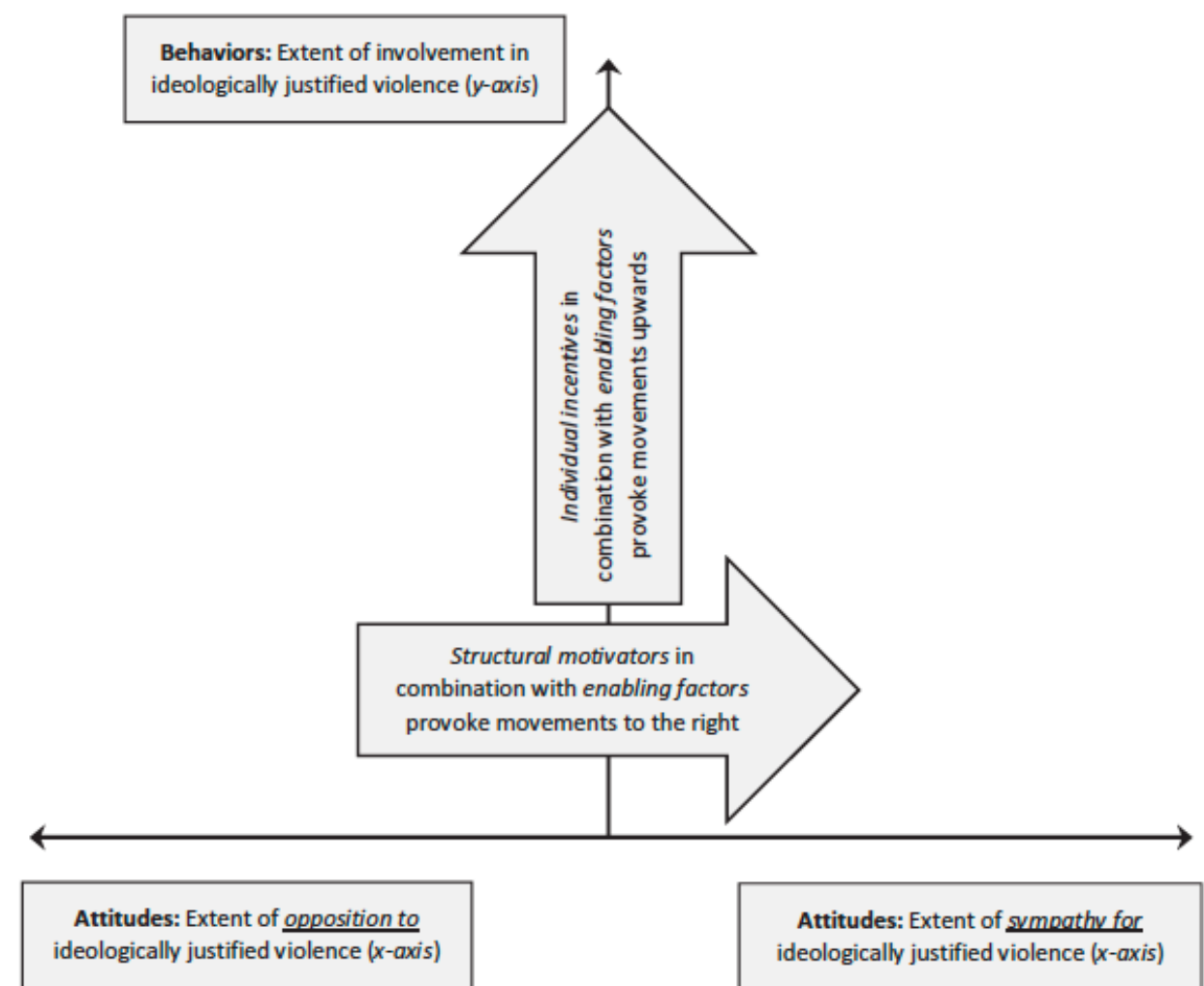


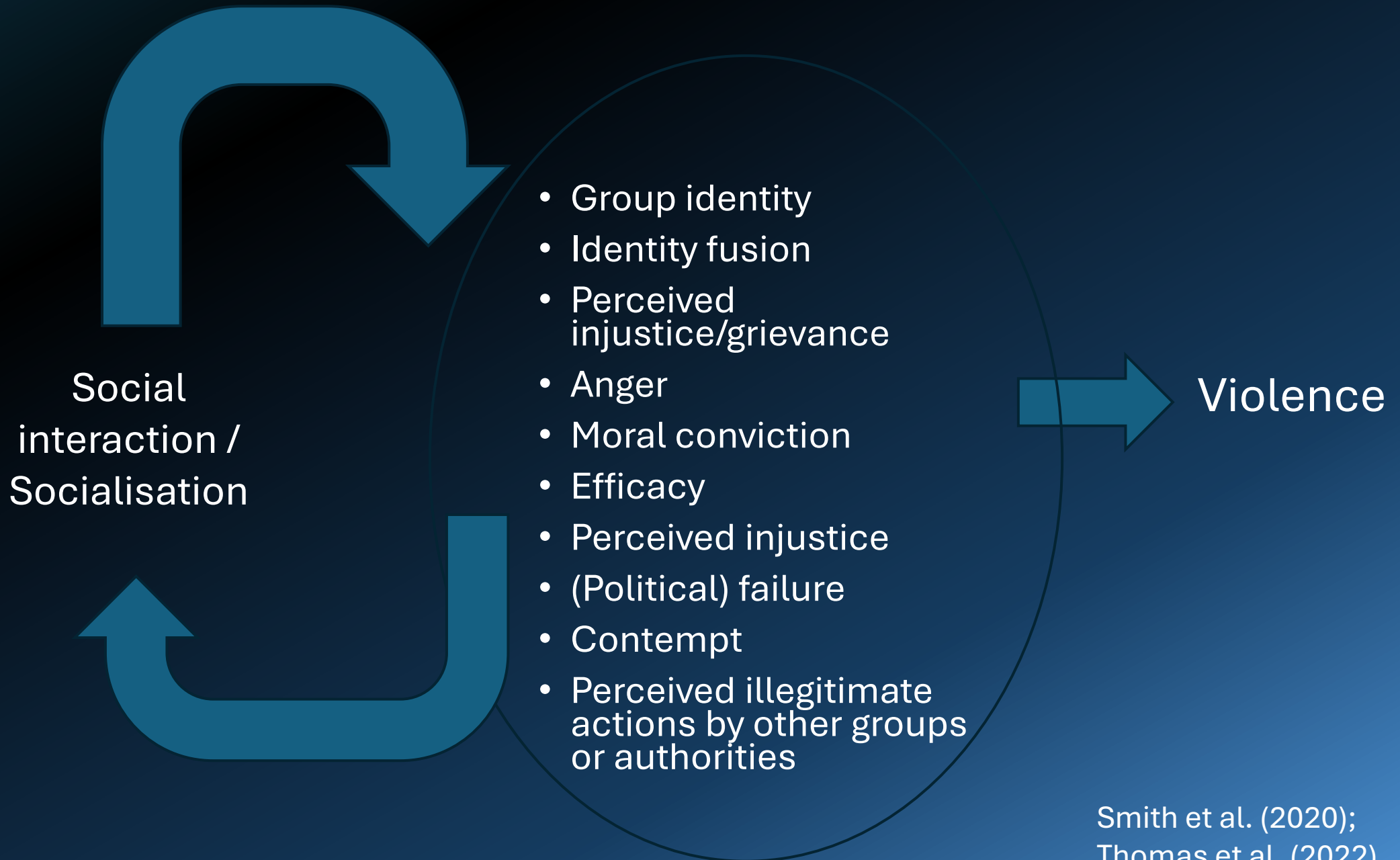
Figure 3. An RCT interpretation of drivers in relation to attitudes and behaviors.

(Group-based) Motivations

- Group identity
- Identity fusion
- Perceived injustice/grievance
- Anger
- Moral conviction
- Efficacy
- Perceived injustice
- (Political) failure
- Contempt
- Perceived illegitimate actions by other groups or authorities

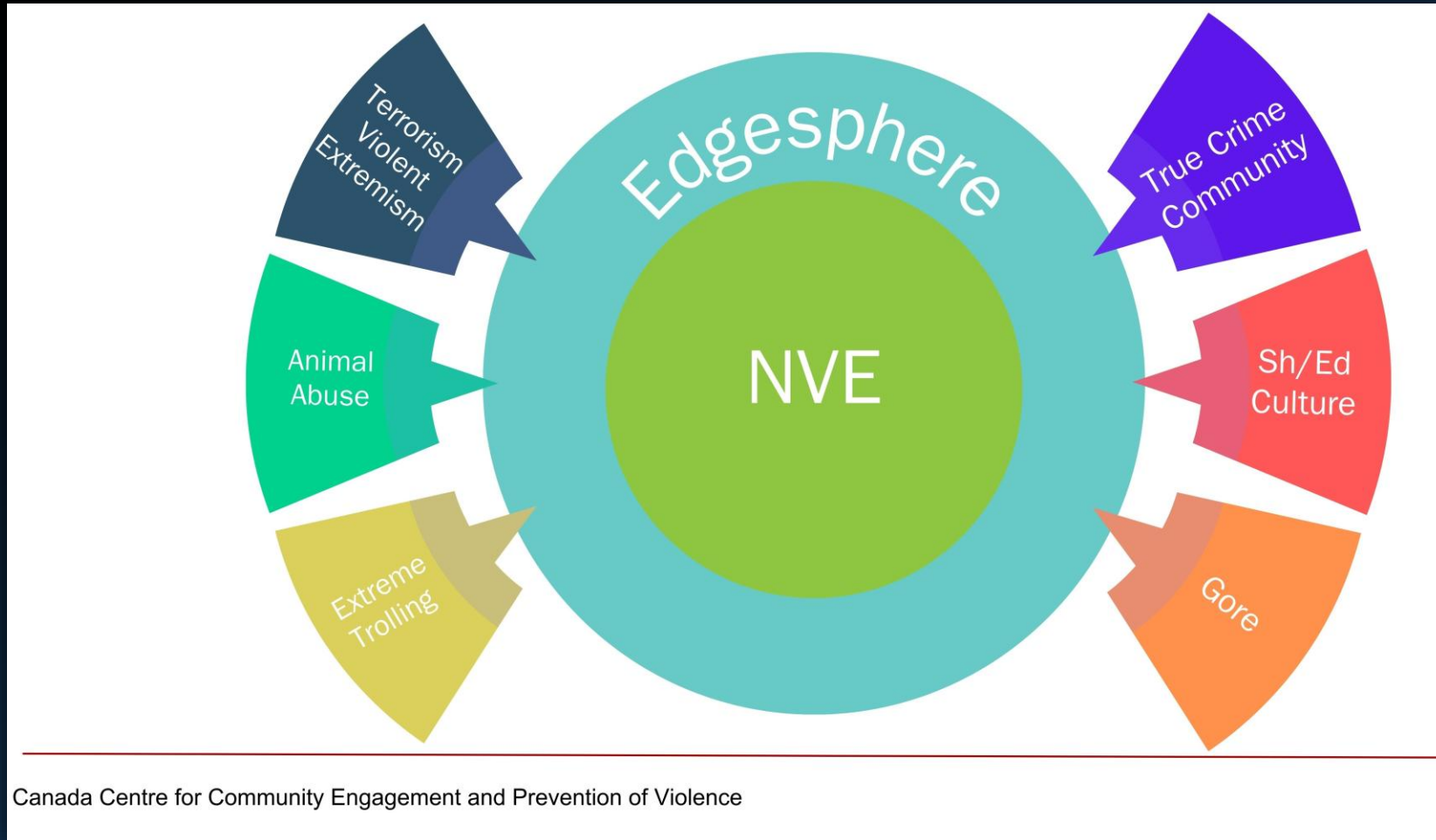


Violence



Smith et al. (2020);
Thomas et al. (2022)

The new reality of online youth radicalisation



Evidence

Work commissioned by the ESRC Network Plus for Analytical Behavioural Science for Security and Defence ([NABS+](#)):

1. **Expert Roundtables:**

- Multi-agency approach and coordination required (Smith & Smith's, 2025, [report](#) and recommendations [here](#))

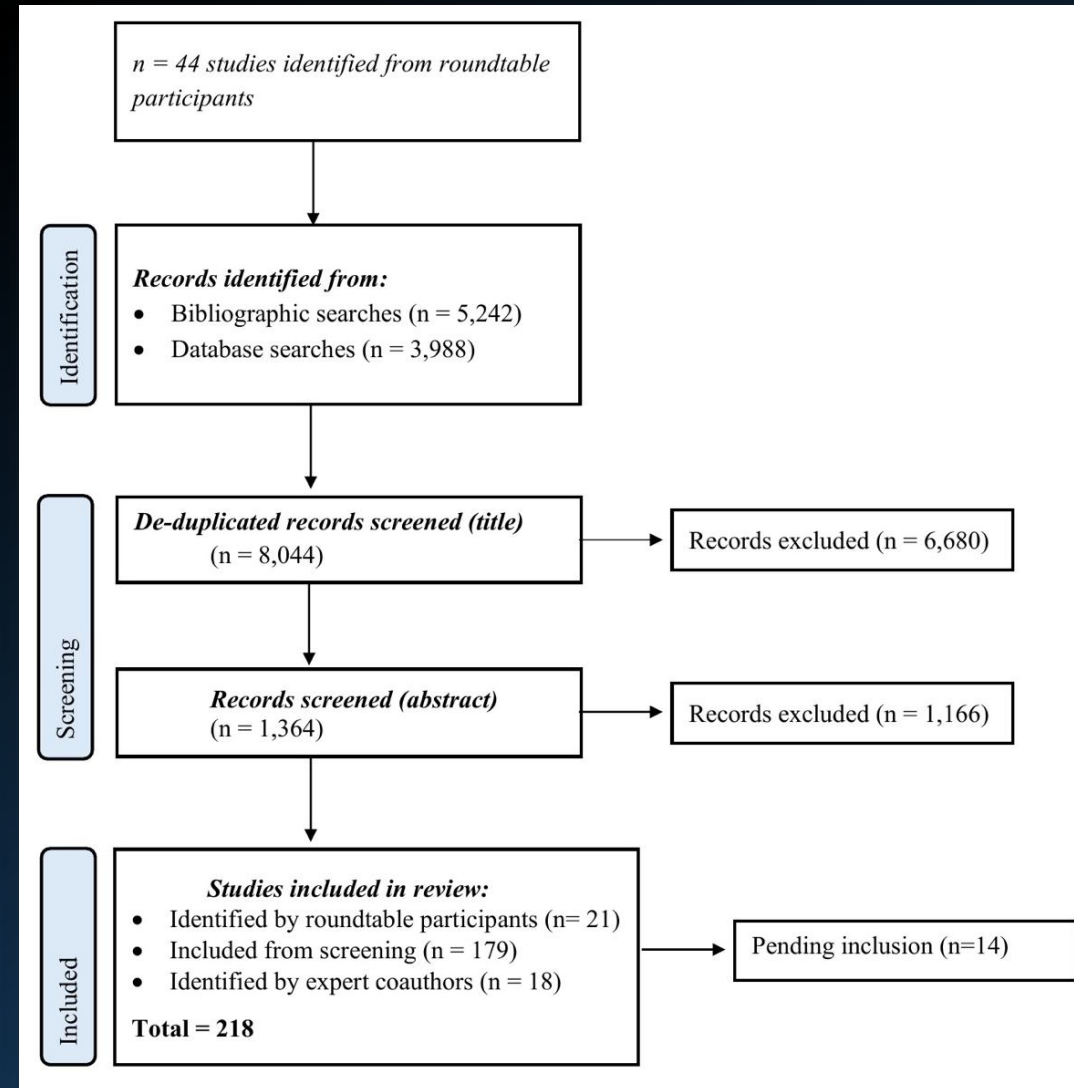
2. **Systematic review** of research on youth exploitation and harm in online networks:

- Public report available [here](#); academic article under peer review available on request

3. **Empirical evidence:**

- The [CONTRA](#) project

Systematic Review (Smith et al., 2026)



<https://crestresearch.ac.uk/resources/youth-exploitation-harm-and-violence-in-online-networks-practitioner-report>

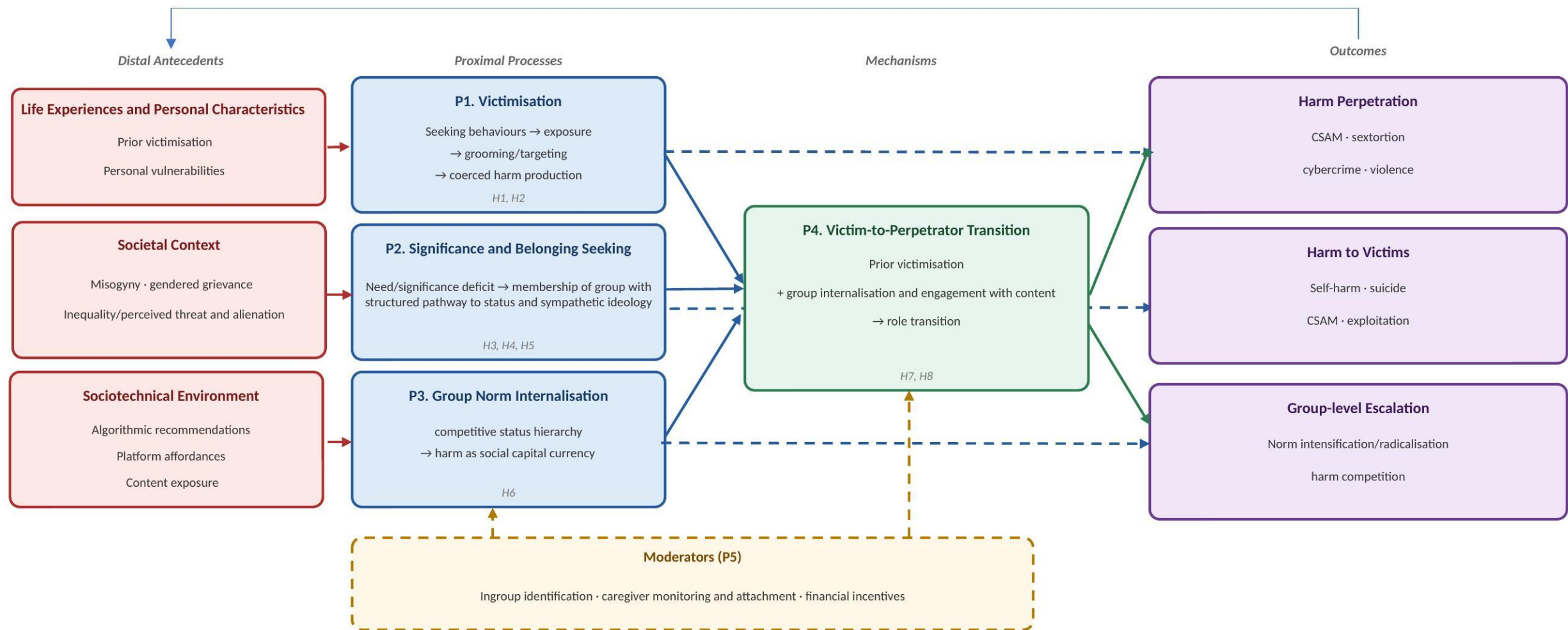
Systematic Review (Smith et al., 2026)

- **Psychological (Personal)** – e.g., history of life experiences, poor caregiver monitoring, perceptions of discrimination (Smith et al., 2026)
- **Social** – social interactions, status-based incentives, norms (Smith et al., 2020)
- **Technical**: – Extreme violent content; social media affordances; recommendation systems; LLMs (Kunst et al. 2026; Smith et al., pre-print)



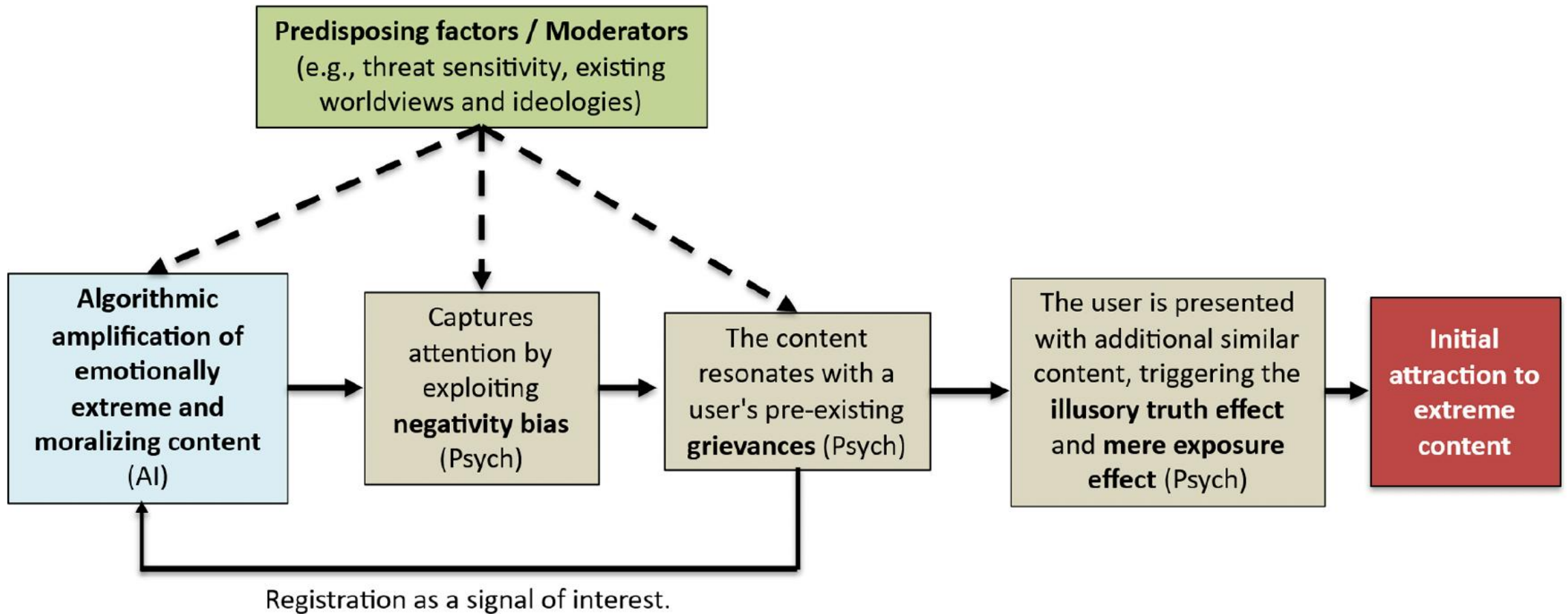
In the absence of ideological drivers or other group-based, sociopolitical motivations, these factors may work in combination to explain the development of motivation for violence.

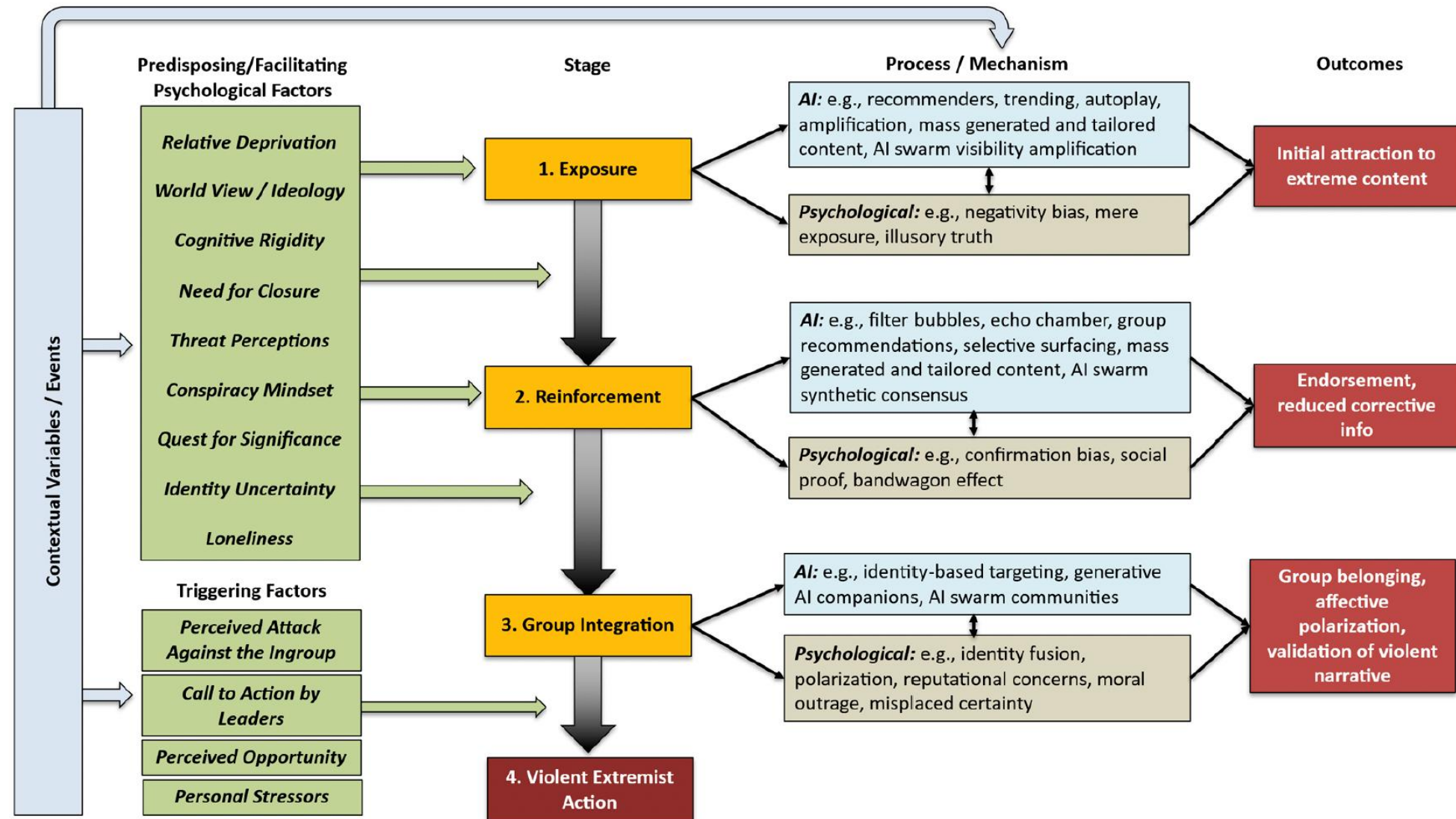
- Capability development (*activation*) is separate and has different signals



Smith et al. (2026)

<https://crestresearch.ac.uk/resources/youth-exploitation-harm-and-violence-in-online-networks-practitioner-report/>





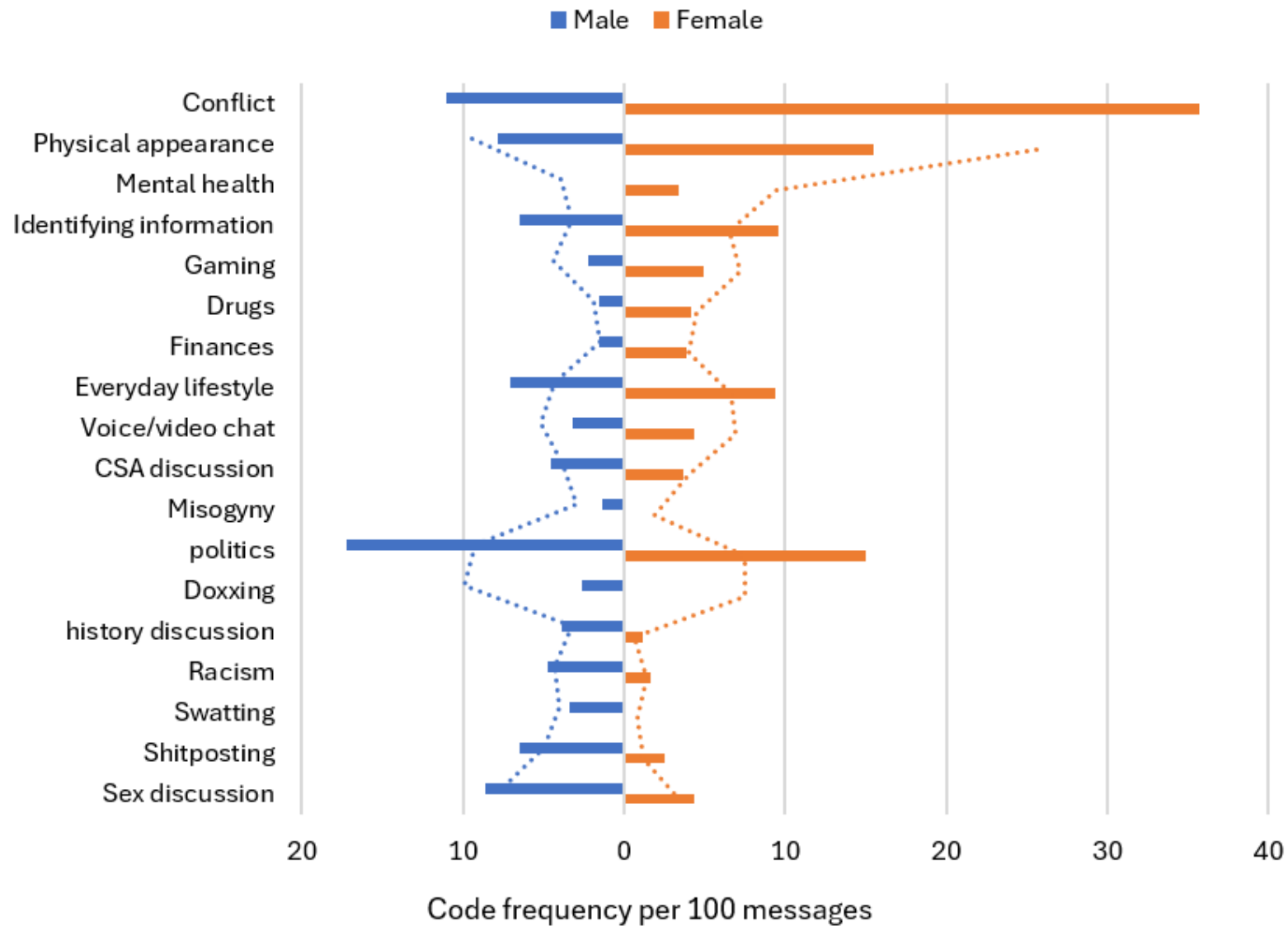
CONTRA

- 1523 users, ~500,000 messages
- October 2023 – March 2024
- Text data – no multimedia content analysed, but high frequency of image, video, audio and streaming media is evident
- Harm is present, but atmosphere is fun – active posting on Christmas and NYE
- Clear (if lenient) moderation policy – reflects a set of values and ethics (albeit contradictory)
- 14 users have received criminal convictions, including the chat admin
- Adjacent chats share 25-50% of their members
- Highly active, disorganised, social messaging channel – a cosmopolitan community with a history and identity
- Links often shared – sometimes to other chats, or channels acting as host for specific extreme content
- Its popularity is borne out of its effectiveness as a social network

Discussion Topics

Derogatory sexual language 15.2%	Conflict and veracity 10.6%	Censorship and moderation 9.3%	Geopolitics 9.2%	Racial politics 6.9%
Commerce and wealth 6.5%	Familial racial slang 5.9%	Informal sexual language 5.7%	Daily life (food) 4.3%	Recreational drug use 3.9%
Family concepts and dynamics 3.3%	Gender politics 2.7%	Death and romance 2.6%	Voice chat 2.4%	Animals 2.1%
Videogames 2.1%	Daily life (sleeping) 2.0%	Cam culture 2.0%	Chat owner discussion 1.8%	Channel-specific discussion 1.7%

Gendered dissociation of message content among influential users



What can research on online mobilisation tell us about the point at which grievance, violent thinking or online expression becomes more closely associated with real-world action?

What psychological, behavioural or social markers most reliably indicate escalation from grievance or fixation towards preparation for violence?

Types of behaviours that might signal risk to society:

- (1) Information gathering; tactical planning; operations (e.g., Schumann et al., 2025)
- (2) Social side – joining and engaging in risky networks/groups (as per CONTRA); evidence of “fusing” with the group (Ebner et al., 2024).
- (3) Group-specific behaviours that reflect conformity to norms of violent groups (as per CONTRA).

BUT: Attitudes are a poor predictor of behaviour; the attitude:behaviour relationship is conditional (Glasman & Albarracin, 2006)

- Key condition is efficacy
- Specificity problem: lots of people have similar radical attitudes, low base rate of offending
- To ascertain risk, research on online mobilisation using a matched-samples design suggests that focusing on signals of activation is useful (Brown et al., 2026; Wolfowicz et al., 2021)

Evidence for association between online behaviours and the perpetration of violence

Brown, Smith, et al. (2026):

Matched case-control design to compare the online behaviour of 26 convicted right-wing extremists with a control group of 48 right wing extremists without convictions.

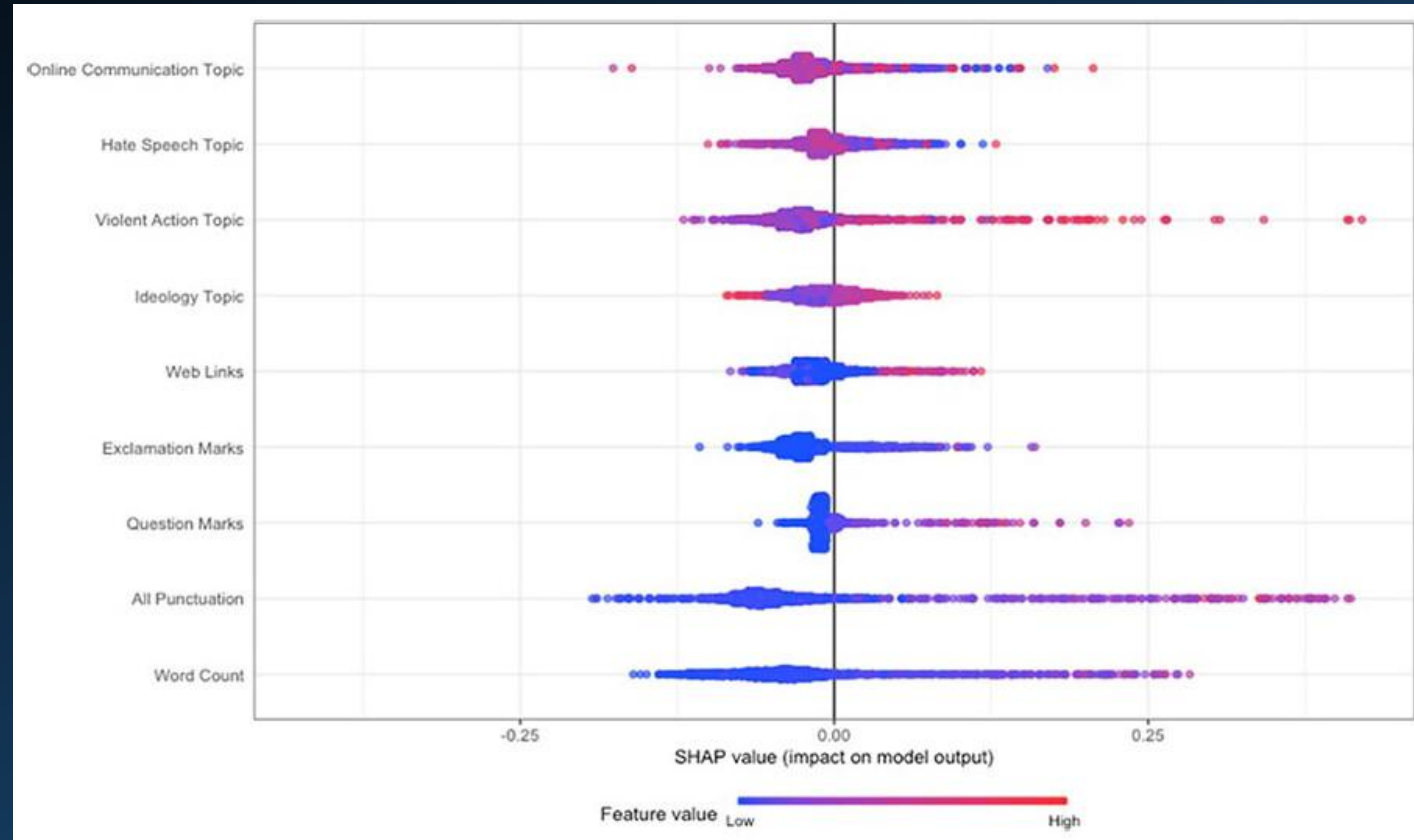
Actual online behaviour, including leakage of behaviour

- 3 convicted of a violent offence involving the murder of one or more people,
- 7 convicted of violent assault,
- 2 convicted of plotting an act of terrorism,
- 7 convicted of possession of terrorism related materials or weapons,
- 3 convicted of membership of a terrorist organisation,
- 3 convicted of making violent threats,
- 3 individuals convicted of distributing harmful content

Evidence for association between online behaviours and the perpetration of violence

Brown, Smith, et al. (2026):

- **Hate/ideological talk** didn't predict conviction status
 - Increases in content relating to hate speech and ideology were related to decreases in the likelihood that a post was authored by a subsequently convicted user.
- **Talk of action, efficacy and logistics** did predict conviction status
 - Content relating to violent action, operational planning, and logistics was most important in distinguishing between convicted and non-convicted users and were more predictive than hate speech or ideological content of future violence.



Evidence for association between Facebook behaviours and the perpetration of terrorism

Wolfowicz et al. (2021): Matched case-control design to compare the Facebook activities and interactions of 48 terrorists in the 100 days prior to their attack with a 2:1 control group of non-violent radicals.

Table 2

Descriptive statistics.

Variable	Terrorists (N = 48)	Non-violent (N = 96)	<i>T</i>	<i>U</i> (Standardized)
Differential associations and its conditioners				
Differential associations with terrorists	0.542 (SD = 0.504)	0.219 (SD = 0.416)	3.837***	3.880***
Network size (Computed)	478.104 (SD = 214.673)	528.083 (SD = 270.561)	-1.116	.199
Network size (Missing) ¹	518.2 (SD = 273.864)	573.533 (SD = 368.005)	-.657	.392
Posts/day (Frequency)	0.555 (SD = 0.795)	0.469 (SD = 0.442)	0.696	-1.344
Duration	38.688 (SD = 20.886)	34.365 (SD = 17.685)	1.300	1.134
Definitions				
Radical ratio	0.696 (SD = 0.397)	0.578 (SD = 0.377)	1.738†	1.804†
Differential reinforcement				
Likes/post	45.001 (SD = 47.136)	44.037 (SD = 36.296)	0.136	-.687
Comments/post	7.538 (SD = 6.813)	9.110 (SD = 9.167)	-1.051	-.161
Shares/post	0.469 (SD = 0.729)	0.156 (SD = 0.326)	2.834**	3.383***
Imitation (post type)				
Text posts (%)	17.938 (SD = 23.089)	31.271 (SD = 22.089)	-3.363**	-3.907***
Shared posts (%)	32.792 (SD = 32.854)	15.271 (SD = 20.637)	3.377***	2.556*
Picture posts (%)	45.083 (SD = 33.285)	45.577 (SD = 26.517)	-0.090	-.352
Video posts (%)	4.20 (SD = .121)	8.00 (SD = .121)	-1.798†	-2.835**

***<0.001, **<0.01, *<0.05, †<0.10.

¹ 25% missing for the terrorist group and 68% missing for the non-violent radicals group.

Summary of Implications as Relevant to Inquiry

- The risk an individual poses needs to be assessed with respect to the different indicators of motivation and capability development in (1) self or (2) others.
- A shared multi-agency risk-assessment tool should include online behaviours, delineated by those that facilitate motivation and activation. Browsing/search history, app usage, gaming, and social media usage needs to be audited as standard and risk assessments repeated over time.
- There should be a further ability to restrict or monitor access to the internet if a significant threshold is passed concerning the risk they pose to others, including restriction of the ability to delete accounts and search history.
- Social media, online games, chat bots, and search engines should have content controls on as a default “out of the box” standard (an opt-out rather than opt-in approach)
- Systems and processes should be reconfigured to remove the ideology requirement in favour of an updated understanding of (all) the processes that lead to a fixation on, or the believed desirability of, violence.

Summary of Implications as Relevant to Inquiry

Regarding the new reality of online youth radicalisation, there is a need for:

- Consensus on a multiagency definition to facilitate operational response
- Strategy to deconflict and clarify operational response.
- Alerts to notify frontline staff, parents, and carers of the risks and signs of risk.
- Development of educational and training material for frontline staff, and for parents and carers.
- Development of mechanism for sharing new insights/standard operating procedures across agencies

(Smith & Smith, 2025)

“The implications of this integrative picture for practice are not trivial: they suggest that effective responses will need to address the sociotechnical environments that generate these conditions [...], rather than focusing exclusively on the attitudes or individual-level characteristics of those who become enmeshed in the networks,” (Smith et al., 2026, p. 29).

Professor Laura Smith

Contact: l.g.e.smith@bath.ac.uk

Key Readings & Evidence*

- *Brown, O., Smith, L. G. E., Davidson, B. I., Racek, D., & Joinson, A. (2026). Online Signals of Extremist Mobilization. *Personality and Social Psychology Bulletin*, 52(1), 70–89. <https://doi.org/10.1177/01461672241266866>
- *Ebner, J., Kavanagh, C., & Whitehouse, H. (2024). Assessing Violence Risk among Far-Right Extremists: A New Role for Natural Language Processing. *Terrorism and Political Violence*, 36(7), 944–961. <https://doi.org/10.1080/09546553.2023.2236222>
- *Glasman, L. R., & Albarracín, D. (2006). Forming Attitudes That Predict Future Behavior: A Meta-Analysis of the Attitude-Behavior Relation. *Psychological Bulletin*, 132(5), 778–822. <https://doi.org/10.1037/0033-2909.132.5.778>
- Khalil, J., Horgan, J., & Zeuthen, M. (2019). The Attitudes-Behaviors Corrective (ABC) Model of Violent Extremism. *Terrorism and Political Violence*, 1-26. <https://doi.org/10.1080/09546553.2019.1699793>
- Khalil, J., Horgan, J., & Zeuthen, M. (2020). The ABC Model: Clarifications and Elaborations. *Terrorism and Political Violence*, 1-8. <https://doi.org/10.1080/09546553.2020.1776703>
- Kunst, J. R., Obaidi, M., Gollwitzer, A., Brandtzæg, P. B., Hinrichs, Y., Saini, N., & Schroeder, D. T. (2026). Intelligent Systems, Vulnerable Minds: A Framework for Radicalization to Violence in the Age of AI. *Personality and Social Psychology Review*, 30(3), 395–426. <https://doi.org/10.1177/10888683261430089>
- McCauley, C., & Moskaleiko, S. (2017, Apr). Understanding political radicalization: The two-pyramids model. *American Psychologist*, 72(3), 205-216. <https://doi.org/10.1037/amp0000062>
- Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist*, 60(2), 161–169. <https://doi.org/10.1037/0003-066X.60.2.161>
- *Schumann, S., Kenyon, J., & Binder, J. (2025). Identifying distinct types of internet use that predict the likelihood of planning or committing a terrorist attack: Findings from an analysis of individuals convicted on terrorism(-related) charges in England and Wales. *Computers in Human Behavior*, 168, 108646. <https://doi.org/https://doi.org/10.1016/j.chb.2025.108646>
- Smith, L. G. E., Blackwood, L., & Thomas, E. F. (2020). The Need to Refocus on the Group as the Site of Radicalization. *Perspectives on Psychological Science*, 15(2), 327-352. <https://doi.org/10.1177/1745691619885870>
- *Smith, L. G. E., Bocheva, D., Brown, O., Lowery, C., Tarpay, H., Torgbi, M., Worajitwannakul, S., & Tayyar Madabushi, H. (pre-print). *Large Language Models May Promote Radicalization Via Exposure To Extremist Content And Politicized Communities*. https://doi.org/https://doi.org/10.31234/osf.io/my3gw_v1
- Smith, W., & Smith, L. G. E. (2025). *Youth Radicalisation and Harm in Online Networks*. NABS+ Findings Report, Expert Roundtables 1 and 2. https://crestresearch.ac.uk/site/assets/files/5641/online_harms_com_nabs_findings_report.pdf
- *Smith, L. G. E., Squires, C., Binder, J., Brown, O., Busher, J., Hamilton-Giachritsis, C., Hughes, N., Kloess, J., Lizzio-Wilson, M., & Smith, W. (2026). *How social and behavioural science insights can support multiagency responses to youth exploitation, harm, and violence in online networks*. Manuscript under review.
- Thomas, E. F., Duncan, L., McGarty, C., Louis, W. R., & Smith, L. G. E. (2022). MOBILISE: A Higher-Order Integration of Collective Action Research to Address Global Challenges. *Political Psychology*, 43(S1), 107–164. <https://doi.org/https://doi.org/10.1111/pops.12811>
- *Wolfowicz, M., Perry, S., Hasisi, B., & Weisburd, D. (2021). Faces of radicalism: Differentiating between violent and non-violent radicals by their social media profiles. *Computers in Human Behavior*, 116, 106646. <https://doi.org/https://doi.org/10.1016/j.chb.2020.106646>