How do game design principles parallel extremist recruitment mechanics, and what practical tools can be developed to address the exploitation of gaming communities for radicalisation?

Dr Simon McCallum, Victoria University of Wellington

This report synthesizes international research, case analysis, and engagement with the New Zealand gaming industry to address this question.

Source Material and Methodology

The research was undertaken using a mixed-methods approach based on:

- **Literature Review:** An examination of international research on the gaming-extremism nexus, drawing on key publications from the Royal United Services Institute Extremism and Gaming Research Network (RUSI EGRN), the Global Project Against Hate and Extremism (GPAHE), the International Centre for Counter-Terrorism (ICCT), and the *Frontiers* journal series on identity fusion in gaming cultures.
- **Case Analysis:** A deep dive into the September 2025 Charlie Kirk assassination attempt, which demonstrated an emerging pattern of violence based on Mixed, Unclear, and Unstable (MUU) ideology, and was saturated with gaming references and irony. The analysis reinforced the distinction that gaming communities, not games themselves, provide the social infrastructure for individuals to act upon extremist ideas.
- **Engagement with the NZ gaming industry**, Conference workshops and Executive Board Meetings.

Key Findings

The research confirmed that online gaming is a significant radicalisation pathway.

International Exposure and Toxicity:
Quantitative findings from a RUSI EGRN study of over 2,200 gamers across seven countries confirmed a high rate of exposure: one-third of gamers reported exposure to extremist content within gaming spaces, and one-quarter reported encountering active recruitment attempts. Furthermore, the gaming environment has persistent toxicity, with more than 50 percent of female gamers experiencing gender-based harassment, and nearly one-third of all gamers perceiving toxicity as a normalized feature of their communities. When the proportion of toxic participants exceeds approximately 10%, cooperative behaviour in a community collapses. Current rates are about 2-3% for functioning game communities. Enough so that most players are exposed to issues, but not to destroy the community.

Economic Consequences:
This toxicity poses a direct material commercial risk to the NZ gaming sector, which generated NZD 759.57 million in revenue in 2024/2025 and is projected to exceed NZD 1 billion in 2026. The RUSI data quantified this damage: 60 percent of gamers reported spending less money, 60 percent reported quitting communities, and 70 percent reported avoiding interactions entirely when exposed to toxicity. For the predominantly small, community-dependent NZ game studios, this attrition can be existential.

Structural Parallels and Vulnerability:
A central finding is that extremist recruitment tactics share structural parallels with game design mechanics because both systems exploit the same core human psychological needs, as defined by **Self-Determination Theory (SDT):** Autonomy (choice/agency),

Competence (mastery/contribution), and Relatedness (belonging/connection). The mechanics are transferable: creating belonging (game mechanic) parallels welcoming vulnerable individuals into an in-group (recruitment tactic), and progression mechanics parallel radicalisation pipeline stages. Unfulfilled psychological needs create vulnerability; for example, a lack of genuine belonging makes an individual susceptible to groups offering false community. The radicalisation process is modelled as a four-stage pipeline: (1) Desensitisation; (2) In-group identity formation; (3) Ideological framing; and (4) Identity fusion and action.

New Zealand Context and Gaps:
While New Zealand has experience with internet-native extremist violence (Christchurch mosque attacks, gamification of the violence), enforcement of online regulation is difficult due to the ephemeral and sub culturally encoded nature of content. The NZ gaming sector identified a gap in purpose-built, research-informed moderation tools for small-to-medium developers that manage community safety and violent extremism prevention simultaneously. Existing educational resources in schools also have gaps, addressing safety as a content or behaviour problem, but failing to equip students with the **psychological needs literacy** needed to understand and resist structural manipulation mechanics.

Key Recommendations

The report recommends the development and deployment of two integrated deliverables: a bot to help manage online chat rooms; and age related training in schools using game design literacy to teach manipulation techniques so students are familiar with the approaches that are used to manipulate them and lure them in. The proposed tools align with the three pillars of the DPMC Counter-Terrorism and Violent Extremism Strategy (December 2025): Identify, Understand, and Disrupt Threats; Prevent and Reduce Radicalisation Risk; and Protect People, Places, and Infrastructure.

1. PvEbot: Community Safety Tool Design
The report recommends consideration of the trialling of **PvEbot** for Discord, in conjunction with the New Zealand Game Developers Association (NZGDA). A prototype of PvE bot has been developed. This tool is designed to fill the moderation gap for small-to-medium studios.

- **Detection Architecture:** PvEbot uses seven rule-based detection analysers (Keyword, Pattern, Behavioural, Contextual, Post-ironic, Recruitment, and Identity fusion) that do not rely on generative AI, addressing developer concerns about sending community data to overseas providers. The Post-ironic violence analyser is novel, designed to detect MUU-pattern content where violence is normalized through irony and gaming language.
- **Graduated Response System:** It employs a seven-level (Level 0 to Level 6) graduated response system, where only the lowest levels (L0-L2) are automated, and higher-level interventions (L3-L6, including Mute, Kick, and Ban) require human moderator approval. This system emphasizes interventions proportionate to behavior severity, avoiding punitive bans that can accelerate radicalisation.
- **Rehabilitation Pathways:** Every ban category (except illegal activity) includes a structured rehabilitation pathway, reflecting the principle that de-radicalisation requires relationships. The most intensive is the **Mentored Rehabilitation** pathway for extremist content (60-180 days), involving trained community volunteers and guided conversations to address the individual's unmet SDT needs. The design is restorative, grounded in the Te Ao Māori value of *Manaaki* (caring/hospitality).

- **Deployment:** PvEbot is designed for deployment on New Zealand-based cloud infrastructure (e.g., Catalyst Cloud) to ensure data sovereignty and offer scalable, low-cost protection for the entire sector (a shared instance costing approximately $45 per month).

2. Age-Differentiated Educational Content

The report recommends the introduction of age-specific educational content for New Zealand schools (Years 1 to 13) on game design literacy leading to understanding manipulation techniques.

- **Focus:** A draft of this material has been developed which teaches students **game design literacy** so they can understand and resist the structural mechanics of manipulation, explicitly connecting bullying dynamics, an entry-point behaviour, to the pathways of radicalisation.
- **Curriculum and Framework:** The content of the draft material aligns with the Te Mataiaho curriculum refresh, particularly the Civics and Society strand. It is grounded in a social constructivism pedagogy (activity first, debrief second) and uses the psychological frameworks of SDT and Te Ao Māori values, such as *Mana Motuhake* (Autonomy) and *Whānaungatanga* (Relatedness), as an integrated framework for building resilience.
- **Differentiation:** The draft content is differentiated across five age bands. Senior students (Years 9-13) learn to use SDT as an analytical framework to deconstruct game mechanics and recruitment tactics, engage with concepts like identity fusion, and participate in policy brief assignments. The content is designed to complement existing Netsafe and Keep It Real Online resources and includes alignment with NCEA assessment standards.

The report concludes by recommending that DPMC consider supporting the further development and testing of PvEbot and that the Ministry of Education should consider teaching students age differentiated game design literacy.

Notes

A full version of the research report is available from DPMC.