The Identity–Extremism Nexus in Virtual Groups

The Impact of Online Group Alignment on Radicalisation Towards Violence

Julia Ebner

St John's College
Institute of Cognitive and Evolutionary Anthropology
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Supervised by Prof. Harvey Whitehouse and Dr. Christopher Kavanagh

April 2023

Word Count: 70,283

Table of Contents

ABSTRACT 4

ACKNOWLEDGEMENTS 5

DECLARATIONS 6

ABBREVIATIONS 8

INTRODUCTION 10

- Research Aim and Hypothesis 10
- Defining Extremism, Violent Extremism, and Terrorism 13
- Defining Groups 16
- Theory of Language 18
- Navigating Cyber Prefixes 20
- Research Design 20
- Research Methods 23
- Thesis Outline 24
- Scholarly Contribution 27

CHAPTER 1: SOCIO-PSYCHOLOGICAL RADICALISATION DRIVERS 28

- Psychological Explanations of Radicalisation 28
- Psychological Theories of Radicalisation 31
- Online Radicalisation Models 35
- Identity Fusion: An Emerging Psychological Explanation of Violent Extremism 39
- Measuring Forms of Group Cohesion 43

CHAPTER 2: THESIS METHODOLOGY 47

- Data Mining 49
- Data Storage 54
- Natural Language Processing (NLP) 55
- Digital Ethnographic Research 59
- Ethical Considerations 69

CHAPTER 3: TOWARDS A NEW VIOLENCE RISK ASSESSMENT FRAMEWORK 73

- Terrorism Prediction Models 76
- Using Language to Detect Would-Be Terrorists 83
- Qualitative Comparative Manifesto Analysis 90
- Manifesto Selection 91
- Manifesto Coding 102
- Narrative Coding 104
- Intercoder Reliability (ICR) Test 108
- Narrative Patterns and Linguistic Proxies 111
- Quantitative Manifesto Analysis 131
- Manual Review 133
- Natural Language Processing (NLP) Findings 134
- Limitations of the Framework 138
- Conclusions from the Manifesto Analysis 143

CHAPTER 4: INVESTIGATING VIOLENCE RISK IN EXTREME-RIGHT ONLINE GROUPS 146

- Introduction 146
- Existing Literature 148
- Approach and Methods 150
- Quantitative Results 154
- Qualitative Results 158
- Research Limitations of the Case Study 169

Case Study Conclusions 170

CHAPTER 5: INVESTIGATING VIOLENCE RISK IN ONLINE QANON GROUPS 172

Introduction 172
Existing Literature 175
Approach and Methods 178
Quantitative Results 183
Qualitative Results 185
Research Limitations of the Case Study 193
Case Study Conclusions 196

CHAPTER 6: CONCLUSIONS 199

Implications for Theories on Radicalisation, Violent Extremism, and Terrorism 202
Research Limitations and Follow-Up Projects 204
Practical Implications for Prevention 208
Practical Implications for Deradicalisation 211
Practical Implications for Online Interventions 214
Ethical Limits and Lessons from the Past 219
The Future of Research, Policy, and Practice 224

BIBLIOGRAPHY 230

LIST OF FIGURES AND TABLES 255

APPENDICES 256

Appendix 1: R Codes for Manifesto Analysis 256
Appendix 2: R Codes for Online Group Analysis 260
Appendix 3: Detailed ICR Analysis Results 263
Appendix 4: Critical Values of the Mann-Whitney U (two-tailed testing) 265

Abstract

Can would-be perpetrators of extreme violence be reliably identified by the linguistic traces they unintentionally leave behind in their online communications? Previous research has shown that identity fusion, a visceral feeling of oneness with the group, can motivate violent self-sacrifice when the in-group is threatened. This thesis tests the hypothesis that a higher degree of identify fusion in online groups, coupled with a range of mediating and moderating variables, increases the likelihood of individuals carrying out violent pro-group action. To date, there has been no systematic analysis of the relationship between identity dynamics within virtual communities and radicalisation towards violence. This research project seeks to fill this gap by investigating the relationship between identity fusion in the digital arena and violent extremist activities in the real world. Using a mixed methods approach, the project combines quantitative natural language processing (NLP) analysis with quantitative text analysis and digital ethnographic research. The first part of the thesis presents a new coding scheme that traces the narratives and linguistic markers found across the written statements published by terrorists prior to launching an attack. A total of 4,000 pages by 15 authors, on a spectrum from violent terrorist to non-violent political manifestos, were analysed in a comparative manifesto analysis and tested via intercoder reliability (ICR) analysis. The statistical and ethnographic findings indicate that linguistic proxies for identity fusion and other relevant variables, such as violence-condoning group norms and dehumanising vocabulary, can be reliably identified and are significantly more prevalent in the documents of would-be terrorists. A weighted score, the so-called "Violence Risk Index", was created, drawing on the statistical findings of the manifesto analysis. This new language-based violence risk assessment framework was then applied to eight online groups varying in their degree of extreme ideologies, verbal commitment to violence, and real-world links to terrorist activities. A total of over one million messages were collected from online forums and messaging apps and investigated with a view to determining the scale and nature of violence-predicting narratives and language in each of the groups. The calculated Violence Risk Index for each group mirrored the level of real-world engagement in extreme violence and terrorism, thus providing preliminary evidence that the assessment framework offers an accurate estimation of violence risk associated with online groups. This thesis argues that psychologically grounded linguistic markers are a more reliable predictor of extreme violence than taking violent threats at face value. While this project's findings should be treated with caution, they may contribute to complementing and improving existing early warning systems used by security and intelligence services.

Acknowledgements

First of all, I would like to thank my brilliant supervisors, Harvey Whitehouse and Chris Kavanagh, for their invaluable support, guidance, and feedback throughout this entire DPhil project. They have been the best possible mentors along an inspiring three-year journey through fusion theory, data analysis, and violent extremist content. I am also deeply grateful for the support I received from the Economic and Social Research Council (ESRC) and St John's College throughout the duration of this project. I feel incredibly lucky to have been part of the Grand Union Doctoral Training Partnership (DTP) and the network of outstanding fellow DPhil students from a wide range of disciplines. My gratitude is also directed to my viva voce assessors Ralph Schroeder and Peter Neumann for their in-depth engagement with the thesis, their thoughtful comments and inspiring feedback. I would furthermore like to thank David Zeitlyn and William Kelly for introducing me to scientific debates and methods in digital ethnography, and express my gratitude to Javier Lezaun, David Pratten, Daniel Köhler, and Martha Newson for their feedback and inspiration. I am deeply indebted to Philip Kreißel, an excellent data scientist at the University of Bamberg, HateAid, and Volksverpetzer for his tech expertise, mentoring, and advice in the natural language processing (NLP) coding process. Many thanks also go to the two independent expert coders, Jacob Davey and Graham Macklin, as well as the independent coders from King's College London's Department of War Studies, Ethan Bossuyt, Sarah Cammarata, Jordan T. Chapman, Asha Cornelius, Louis Dean, Aparajita Arun Deshpande, Gaia Gaddi, Claire Gibbons, Emily Glynn, Oliver Hair, Spencer Helsing, Courtney Kay, Aaron Kiesler, Katherine E. Lee, Matteo Natlacen, Shaheer Iqbal, Louisa Ryder, Anke Schlieker, Antonia Sheppard, Olivia Singer, Nidia Sookhoo, Leo Thorncroft, Freya Thorner, and Jennifer Young. Finally, a big thank you goes to my family and friends, in particular my husband and son, who have been by my side and a daily motivating force behind my work.

Declarations

Funding Declaration

This work was generously supported by the UK's Economic and Social Research Council (ESRC) and St John's College at the University of Oxford under the ESRC Grand Union Doctoral Training Partnership (DTP) Studentship.

Disclosure Statement

No potential conflict of interest was reported by the author of this thesis.

Data Availability Statement

Due to the sensitivity of the analysed content, the researcher refrains from publishing any raw datasets. However, all full manifestos and coding sheets can be made available upon request to academics and experts who can provide proof of their affiliation with an independent research institution.

Declaration of Authorship

I hereby confirm that this work is my own, except where otherwise specified. My doctoral project takes the form of an integrated thesis. Several of its chapters have been published, will be published, or are under consideration for publication in peer-reviewed journals and edited volumes. I am responsible for the majority contribution to all publications featured in this integrated thesis, including the conceptional development, analysis, and write-up. In accordance with the Postgraduate Research (PGR) Handbook guidelines, all submitted work was produced during, and based on, research in connection with the DPhil programme. This thesis features chapters based on the following academic papers and book chapters:

- Ebner, Julia, Chris Kavanagh and Harvey Whitehouse, "Is there a language of terrorists: A comparative manifesto analysis", *Studies in Conflict & Terrorism*, August 2022.

- Ebner, Julia, Chris Kavanagh and Harvey Whitehouse, "The QAnon security threat: A linguistic fusion-based violence risk assessment", *Perspectives on Terrorism,* Special Issue, December 2022.

- Ebner, Julia and Harvey Whitehouse, "Identity and extremism: Sorting out the causal pathways to radicalisation and violent self-sacrifice" in Joel Busher, L. Malikki and S. Marsden (Eds.), *Routledge Handbook on Radicalisation and Countering Radicalisation* (Abingdon: Routledge, 2023).

- Ebner, Julia, Chris Kavanagh and Harvey Whitehouse, "Assessing violence risk among far-right extremists: A new role for natural language processing", *Terrorism and Political Violence* – forthcoming 2023.

- Ebner, Julia, Chris Kavanagh and Harvey Whitehouse, "Measuring socio-psychological drivers of extreme violence in text: An alternative linguistic risk assessment model" (submitted to the *Journal of Policing, Intelligence and Counter Terrorism*).

The full articles and book chapter publications will be submitted as Additional Materials. A detailed description of how these chapters were integrated to form part of a coherent doctoral thesis is provided in the Introduction's sub-section "Thesis Outline". In order to avoid excessive repetition of the underlying theoretical frameworks and methods used, the published materials were occasionally adapted for this thesis and chapters were cross-referenced. Other publication-based chapters have been amended for the purpose of the thesis to provide additional methodological details and present interim findings of the analysis.

Abbreviations

ABC	Attitudes-Behaviours Corrective
AI	Artificial Intelligence
alt-right	Alternative Right
alt-tech	Alternative Technology (Platforms)
Anticom	Anti-Communist Action
Antifa	Anti-Fascist
AOIR	Association of Internet Researchers
CASM	Center for the Analysis of Social Media
CBRN	Chemical, Biological, Radiological, and Nuclear
CIA	Central Intelligence Agency
C-REX	Centre for Research on Extremism
CT	Counter-Terrorism
CUREC	Central University Research Ethics Committee
CVE	Countering Violent Extremism
DHS	U.S. Department of Homeland Security
DIFI	Dynamic Identity Fusion Index
DTP	Doctoral Training Partnership
EASA	European Association of Social Anthropologists
ESRC	Economic and Social Research Council
FBI	Federal Bureau of Investigation
GDPR	General Data Protection Regulation
GNET	Global Network on Extremism and Technology
GTD	Global Terrorism Database
GTI	Global Terrorism Index
IBR	Internet-Based Research
ICCT	International Centre for Counter-Terrorism
ICR	Intercoder Reliability
ICSR	International Centre for the Study of Radicalisation
IDPG	Identification with a Psychological Group
Incel	Involuntary Celibate
ISD	Institute for Strategic Dialogue
ISIS	Islamic State of Iraq and Syria
LARP	Live Action Role Play
LDS	Latter Day Saint
LIWC	Linguistic Inquiry and Word Count
LSE	London School of Economics
MUVE	Multi-User Virtual Environment
NCTC	National Counterterrorism Center
NLP	Natural Language Processing
NS	National Socialist
NSDAP	National Socialist German Workers' Party
ODIHR	Office for Democratic Institutions and Human Rights

OII	Oxford Internet Institute
OSCE	Organization for Security and Cooperation in Europe
PERIL	Polarization and Extremism Research and Innovation Lab
PST	Norwegian Police Security Service
RAN	Radicalisation Awareness Network
RMIT	Royal Melbourne Institute of Technology
RUSI	Royal United Services Institute
START	National Consortium for the Study of Terrorism and Responses to Terrorism
TPIMs	Terrorism Prevention and Investigation Measures
UN	United Nations
UNCTED	United Nations Counter-Terrorism Committee Executive Directorate
UNDP	United Nations Development Programme
UNODC	United Nations Office on Drugs and Crime
VR	Virtual Reality
WWG1WGA	Where We Go One We Go All
ZOG	Zionist Owned Government

Introduction

Research Aim and Hypothesis

This research project seeks to explore the impact of extreme forms of group cohesion in virtual spaces on radicalisation towards violence through an interdisciplinary lens, building on existing research from the fields of cognitive anthropology, social psychology, internet studies, and linguistics. To date, there has been no systematic analysis of the relationship between identity transformation processes within digital spaces and radicalisation towards violence. This research project therefore seeks to fill this gap by exploring the impact of digital identity fusion on the risk that members of online groups resorting to forms of real-world violence. Beyond its contribution to filling academic gaps related to the study of online radicalisation and the online–offline extremism nexus, this research project could also provide valuable insights for policymakers, social media companies, and frontline practitioners seeking to understand, prevent, and combat online radicalisation towards violence.

In recent years, both online radicalisation towards violence and campaigns to intimidate politicians, journalists, and political activists have become growing phenomena. Security experts have warned of the spread of violent extremist materials in online forums and encrypted messaging apps (Europol, 2021; Kenyon et al., 2021). New research has found a significant rise in the launching of harassment campaigns against politicians, which has caused many victims to experience high levels of anxiety and even to retreat from their profession (Bjorgo et al., 2022). According to a global study of over 900 journalists conducted in 2020, almost three in four female journalists

face online violence. Mainstream and fringe cyberspaces (so-called "alt-tech" platforms) have turned into hotbeds for viral hate and violence inspiration, leaving security services often overwhelmed with the sheer amount of potential threats to national security (Guhl et al., 2020).

Not everyone who makes explicit threats of violence will translate their words into action. Likewise, not everyone who will commit an act of extreme violence threatens to do so beforehand. Some will even intentionally stretch the boundaries of legality by using the rhetorical weapon of satire (Ebner, 2020, Fielitz & Thurston, 2019), leading intelligence officers into a tricky security-versus-freedom dilemma. Judgement on whether a user who engages in "shitposting" constitutes a risk to violence can be difficult; however, the stakes are high. The intersection of satire and hate has escalated in a new phenomenon of political violence: gamified terrorism (Ebner, 2020). As we have been able to observe the growing relationship between trolling and terrorism, traditional counter-terrorism (CT) mechanisms have become increasingly unhelpful in assessing and dealing with this emerging threat.

The Christchurch attack of March 2019, which left 51 people dead in two mosques in New Zealand, was the first case of gamified terrorism (Schlegel, 2021). Its perpetrator deliberately attempted to blur the lines between fiction and reality. His manifesto and livestreamed footage were filled with insider references of the online gaming community. The first user responding to the 2019 Christchurch attacker's livestream on the image board 8chan wrote: "Is this a LARP [Live Action Role Play]?" (Ebner, 2020). A few months later, a terrorist attacker in Halle, Germany, mirrored the Christchurch methods: he produced a first person shooter livestream of his anti-Semitic

and anti-migrant attack, printed his own three-dimensional (3D) weapons in a reference to what gamers call “weaponscrafting”, and drafted a so-called “achievements” list with scores for each victim he would kill or injure (Lakhani, 2022). This trend has continued to manifest itself in a series of copycat attacks, from the attacks in Poway and El Paso a few months after Christchurch to the 2022 Buffalo shooting (Abbas, 2022).

In the course of the last seven years, I have been invited to many internal briefings and conferences of security and intelligence agencies across the UK, Europe, North America, and Australia. These agencies’ investigations of potential terror threats over that time have proved difficult, partly because they had too much data. A single Europol investigation might contain several million data points that need to be assessed as quickly as possible. Likewise, national intelligence agencies are frequently faced with challenges around prioritisation and resource allocation. In light of the growing phenomenon of gamified radicalisation, many of them have asked me the same question: How can we distinguish between trolls and those actually at risk of committing violence? How do we know when users are just “LARPING” and when they are serious about their violent fantasies?

The aim of this doctoral project was to look beyond explicit threats to violence, a metric that has been traditionally used in the detection of potential terrorists (Council of Europe, 2022; Federal Bureau of Investigation [FBI], 2022; HM Government, 2015). Self-censorship has become common among violent extremists in response to platform removal policies (Calderón et al., 2021). Violent extremist movements have adopted more careful language and even used code words to circumvent detection and pre-empt the closure of their accounts. For example, Calderón et al. (2021) cited code words such

as “skypes, googles, bing, yahoos” as having a hidden hate speech meaning. Although initial efforts have been made to create code-word-resilient detection mechanisms, these tools are still in their testing phase and rely on pattern extraction from hate speech specifically (Calderón et al., 2021). This prompts the question of whether there are other clues in the language of users who are prone to the use of real-world violence that can be used to flag them for investigation. Do perpetrators of extreme forms of violence inadvertently give away their intentions in what they say? In this thesis, I argue that the psychological drivers of political violence are revealed in language even when strategic rhetoric decisions are made to conceal violent intentions. I will present a new framework that helps to better track socio-psychological patterns observed in terrorist communications.

The purpose of the research project is to contribute to preventing and disrupting terrorism and violent extremism. Furthering our understanding of the dynamics that fuel radicalisation towards violence in the digital space is necessary to develop effective prediction, prevention, and intervention mechanisms. The complexity of the challenge requires well-designed research projects that draw on existing frameworks and measures introduced to study virtual spaces, forms of group alignment, and radicalisation dynamics.

Defining Extremism, Violent Extremism, and Terrorism

Definitions of extremism, violent extremism, and terrorism radically differ across different geographies and research institutions. The UK government defines extremism as “vocal or active opposition to fundamental values, including democracy, the rule of

law, individual liberty and the mutual respect and tolerance of different faiths and beliefs, as well as calls for the death of United Kingdom armed forces at home or abroad" (HM Government, 2011, p. 34). According to the Swedish government's definition, a violent extremist is someone who is "deemed repeatedly to have displayed behaviour that does not just accept the use of violence but also supports or exercises ideologically motivated violence to promote something" (Government Offices of Sweden, 2011, p. 9). The US Department of Justice (2014, p. 1) describes violent extremism as "encouraging, condoning, justifying, or supporting the commission of a violent act to achieve political, ideological, religious, social, or economic goals".

This lack of an international framework for terrorism and violent extremism has been at the core of ongoing debates in global policy networks (United Nations [UN], 2018). Less democratic countries, such as China, have exploited notions of terrorism and violent extremism to clamp down on political dissidents and ethnic minorities (Maizland, 2019; Roberts, 2018). The UN and the Organization for Security and Cooperation in Europe (OSCE) have therefore called for procedures to ensure national terrorist designations comply with international human rights (OSCE/Office for Democratic Institutions and Human Rights [ODIHR], 2018; United Nations Counter-Terrorism Committee Executive Directorate [UNCTED], 2016). In May 2019, members of the United Nations Security Council suggested updating the UN lists of designated terrorist organisations, which currently name groups affiliated to the Islamic State of Iraq and Syria (ISIS) and Al-Qaeda but do not include other types of violent extremist organisations (United Nations Security Council, 2019).

Many academics and policy institutions distinguish between ideological extremism (also referred to as cognitive extremism) and violent extremism. According to a research paper published by the International Centre for the Study of Radicalisation (ICSR) at King's College, cognitive extremism is the "process through which an individual adopts ideas that are severely at odds with those of the mainstream, refutes the legitimacy of the existing social order, and seeks to replace it with a new structure based on a completely different belief system" (Vidino & Brandon, 2012, p.9). Violent extremism, as the term suggests, adds a further layer of the willingness to employ violence to advance the beliefs embraced in ideological extremism. This doctoral project differentiates between identity-driven types of extremism, namely ideological fixation on extreme ideologies versus violent pro-group behaviour, which can be compared to ideological extremism versus violent extremism. The term "radicalisation" is used in the context of this thesis to describe the process of radicalisation towards cognitive or violent extremism. Different psychological definitions and explanations of radicalisation will be explored in Chapter 1.

Despite the absence of an international consensus on defining terrorism, violent extremism, and extremism, many efforts have been made to find ideologically agnostic approaches to defining varying degrees of extremism. In his influential book *Extremism*, J. M. Berger maps out five crisis narratives that are characteristic of extremists across the ideological spectrum: conspiracy; impurity; dystopia; existential threat; and apocalypse (Berger, 2018, pp. 82–83). The Institute for Strategic Dialogue (ISD) defines extremism, regardless of the political colour or ideological shape it may take, as "the advocacy of a system of belief that claims the superiority and dominance of one identity-

based ‘in-group’ over all ‘out-groups.’ It advances a dehumanising ‘othering’ mind-set incompatible with pluralism and universal human rights” (ISD, 2022). It is important to note that definitions of extremism vary across time and space. Someone denounced as an extremist in the US of the 1960s or in modern-day China might not be labelled an extremist in most of today’s liberal democracies. For instance, the Alabama clergymen labelled Martin Luther King Jr. an “extremist” during the US Civil Rights Movement (University of Groningen, 2013). However, today’s understanding of the term in liberal democratic countries would neither link his ideas nor his behaviours to extremism.

This thesis applies the term “ideological extremism” based on the ISD’s extremism definition. The term “self-sacrificial violence” is used in this thesis to describe violent pro-group behaviours that entail risk to life and limb for the perpetrators, regardless of whether they subscribe to ideological extreme beliefs or not. Self-sacrificial violence may therefore include acts of terrorism as a form of homicidal self-sacrifice and other expressions of violent out-group hostility (Whitehouse, 2018).

Defining Groups

There is no cross-disciplinary consensus on the definition of groups. While group psychologists tend to draw on social identity theory to distinguish between in- and out-groups (Islam, 2014; Tajfel & Turner, 1997), scholars in the field of internet studies often focus more on spatial, temporal, and socio-topical approaches when defining online groups or virtual communities (Martin-Borregon et al., 2014). In the terrorism studies arena, the internet has challenged traditional definitions of “groups”, as it has given rise to post-organisational forms of violent extremism. Notably, online image boards and

forums, such as 4chan and Reddit, have facilitated the joint action of loosely connected networks of users (Davey et al., 2021). While, traditionally, most extremist groups have relied on centralised, hierarchical, and locally rooted structures, today's movements tend to be characterised by highly fluid and ambiguous forms of online membership based on shared culture and ideology (Newhouse, 2021). Can these virtual collectives without clear boundaries still be described as a group? Where should we draw the line when defining an extremist group?

As this thesis specifically explores identity fusion and other psychological metrics in different online contexts, its definition of "group" is rooted in a social psychological understanding of the term. Group psychologists tend to distinguish between relational and categorical group ties (Brewer, 2001; Swann et al., 2012), which are the two most fundamental forms of group alignment of relevance to this thesis. Identity fusion theorists have suggested that relational ties form the basis for local identity fusion via imagistic experiences based on shared episodic memories. In contrast, categorical ties, which form the basis for group identification, arise through routinisation in the doctrinal mode based on shared semantic memories. However, when one fuses with a categorical group (e.g. all Muslims, or fellow countrymen), it produces a kind of hybrid of identification and fusion that has been described as extended fusion (Whitehouse, 2018; Whitehouse & Lanman, 2014).

Hence, in this thesis, a "group" is not defined by the spaces or communication modes through which it operates but rather by the types of bonds between its members. This wide definition of "online groups" therefore includes both tightly organised movements as well as loose networks or collectives that have a shared understanding

of “us” versus “them” but may be more fluid in their membership. For the purpose of this thesis, the term “online group” may be applied both to small, closed groups on encrypted messaging apps (e.g. an alt-right group on Discord or a conspiracy myth group on Telegram) and less formally organised groups on a forum’s discussion board (e.g. a neo-Nazi forum or a football fan forum).

Theory of Language

In this thesis, I argue that language can serve as a marker for deeply rooted psychological impulses. Although the role of language may be interpreted differently depending on the underlying theory of language (e.g. using a structuralist, functionalist, formalist, or post-structuralist lens), the analysis here focuses on one highly salient dimension of language, which is its ability to provide clues to the type and intensity of a person’s group alignments (Whitehouse & Lanman, 2014). Even though this thesis acknowledges the many functions that language may fulfil (including performing, persuading, and expressing ideas and feelings), this project draws primarily on psychological and cognitive interpretations of language as a medium that can express unconscious psychological inclinations (Bilik et al., 2021; Boyd & Schwartz, 2020; Tomasello, 2014).

Language, this thesis argues, can be the outcome of identity transformation processes but it can also be a tool to build an environment in which in-group identity fusion might occur. Research has shown that language can reveal unconscious and subconscious psychological processes (e.g. Freudian slip; see, for example, Bilik et al., 2021 and Poscheschnik & Crepaldi, 2022) but it can also be used strategically to provoke psychological processes (e.g. propaganda; see, for example, Quaranto and Stanley,

2021). Shared kinship metaphors may be used to create or strengthen the conviction of shared experience among members of a group, which in turn can act as a driver of fusion. This means that kinship language may be observed as the result of identity fusion dynamics in a group, or it may be strategically chosen by group leaders to foster group cohesion. In both cases, higher levels of identity fusion would be expected to be linked to kinship language.

Language analysis methods seeking to extract psychological information from verbal behaviours have evolved in parallel with theories that seek to explain how language is linked to human thoughts, emotions, and behaviours. There has been a growing sophistication in psychological language analysis methods, which have become increasingly objective, systematic, and empirical (Boyd & Schwartz, 2020). Most of today's language analysis theories use the so-called "words of attention" approach as a central pillar, which assumes that word frequencies in written or spoken communication reflect a person's attention to a specific domain, even if this is not explicitly or intentionally expressed. For instance, psycholinguistic analyses may focus on "pronouns" as markers of *who* we pay attention to, while "emotion words" might signal attention to specific affective states (Boyd & Schwartz, 2020). As Chapter 2 will explain in further detail, this thesis combines a word count analysis based on natural language processing (NLP) with a manually supported semantic analysis that contextualises the linguistic markers and analyses their meanings in their wider bodies of text.

Navigating Cyber Prefixes

The terms “online”, “digital”, and “virtual” may all be used to describe groups, processes, or interactions in cyberspace. The use of these prefixes varies across different disciplines, sectors, and organisations, partly reflecting differences in the understanding of their semantics. “Virtual” signals that something is intangible and potentially not real, “digital” reflects on the binary codes consisting of the digits 0 and 1 used in computer systems, and “online” means with connection to the internet (Diplo, 2023). Despite acknowledging these subtle differences, this thesis may occasionally use the terms interchangeably to describe groups or interactions that are all of the above: virtual; digital; and online.

Research Design

The conceptual basis for the project was provided by fusion theory: identity fusion is an extreme form of group alignment whereby an individual’s personal and group identities become functionally equivalent (Swann et al., 2014). In a number of recent studies, the socio-psychological phenomenon of fusion has been shown to drive extreme violence in groups as diverse as Libyan revolutionary battalions (Whitehouse et al., 2014), Indonesian religious fundamentalists (Kavanagh et al., 2020)[1], Cameroonian herders and farmers (Buhrmester et al., 2020), and British and Brazilian football hooligans (Newson, 2019).

[1] This study in the Indonesian context did however find that identification was a stronger predictor of extreme violence than fusion.

Previous research on extreme self-sacrifice conducted in offline settings has found that individuals who show high levels of identification with their in-group are more likely to move towards ideological fixation, whereas high degrees of identity fusion are more likely to lead group members to engage in violent pro-group behaviour, i.e. to fight and die for their group (Whitehouse, 2018). This project aims to test whether the same dynamics apply to online groups. Hence, high identification and low fusion are expected to foster ideological fixation on the virtual group's belief system. Depending on the nature of the ideological doctrines shared in a virtual space, this could take the form of ideological fixation on non-extreme or extreme beliefs. The combinations of high fusion and low identification, as well as of high fusion and high identification, are predicted to strengthen the willingness of group members to engage in violent pro-group behaviour. Low levels both of identification and fusion would indicate a low likelihood of either ideological fixation or violent pro-group behaviour. The proposed relationships between group alignment and behavioural outcome are presented in Figure 1.

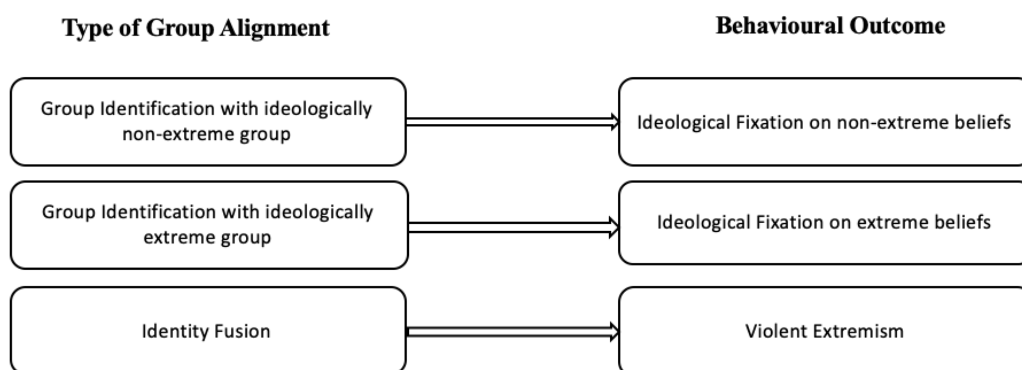


Figure 1: Proposed Relationships Between Group Alignment and Behavioural Outcome

Specifically, the research design is based on the fusion-plus-threat model, an integrated conceptual framework designed by Harvey Whitehouse in 2018. The fusion-plus-threat model aggregates findings from previous studies that provide empirical evidence for the fusion–violence link, as well as the role of perceived existential threat to the in-group in mediating this relationship (Whitehouse, 2018). Taking into account the open peer commentary in response to Whitehouse's model, as well as existing literature on radicalisation pathways to violence, a range of potential mediators between identity fusion and violence were included in the analysis (Whitehouse, 2018). Most notably, the following variables were assessed in this thesis (see Figure 2):

(1) Perceived out-group entitativity (Choi et al., 2018, Lickel, et al., 2006), including the use of "us versus them" narratives (Berger, 2017) and language that insults, demonises, or dehumanises an entire out-group (Giner-Sorolla et al., 2011).

(2) Perceived out-group threat (Whitehouse, 2018), including narratives of an existential threat posed to the in-group (Hirschberger et al., 2016), the belief in a conspiracy of the out-group (Mashuri & Zaduqisti, 2015; Ebner, 2017), and the belief in an inevitable war between the in- and out-group (Louis et al., 2018).

(3) Violence-condoning norms (Louis et al., 2018), including the justification of violence, the glorification of violence via martyrdom narratives or the so-called "warrior mentality" (Louis et al., 2018; Moghaddam, 2018), identification with a violent role model (Cohen et al., 2014; Davey & Ebner, 2019), and the perceived hopelessness of alternative solutions (Thomas & Louis, 2014).

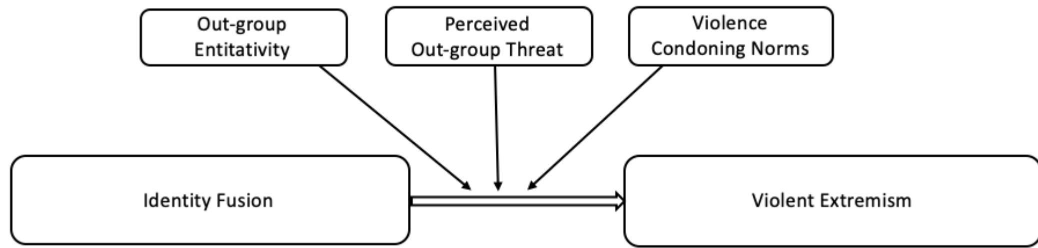


Figure 2: Proposed Relationship Between Fusion and Violence

The hypothesis of this thesis is that high levels of identity fusion are needed to make group members prone to violent pro-group behaviour. Fusion alone, however, might not be sufficient to explain pathways to violence. If there is no perceived threat from and/or belief in an inevitable war with a clearly defined out-group (e.g. expected race war), for example, an attack on behalf of the fused group would seem pointless. Likewise, if other solutions (e.g. political solutions) are regarded as viable alternatives, violence might not be considered necessary. The identification with violence-endorsing group norms (i.e. “warrior mentality”) or violent role models (i.e. “inspirational terrorism” or “copycats”) may also be key mediating and moderating variables in the pathways to violence. This thesis will therefore examine whether the presence of these mediating and moderating factors increases the likelihood of expressed willingness to engage in violent pro-group behaviour, compared to fusion on its own.

Research Methods

To test the overall research hypothesis, a mixed methods approach was used, which encompassed quantitative NLP analysis, qualitative text analysis, and digital

ethnographic research. The analysis was conducted in three steps. In a first step, a curated selection of manifestos, ranging from violent terrorist to non-violent ideologically moderate, were analysed qualitatively and quantitatively. Recurring narratives and their linguistic markers were distilled from the writings of authors who then committed extreme forms of violence and compared to the patterns found in ideologically extreme or moderate manifestos that were not followed by violent activity by their authors. The aim of this part of the thesis was to measure the occurrence of the above-mentioned variables of interest in violent terrorist manifestos, as well as to assess whether the identified patterns were in line with the fusion-plus-threat model and existing literature on violent extremism. Based on this comparative manifesto analysis, a text-based violence risk assessment framework grounded in fusion theory was designed.

The next step was to apply this new framework to a range of extremist and non-extremist online forums across the violence spectrum to further refine and test the framework. Finally, the last step was to compare and contrast the results from all assessed groups. While the results of the NLP analysis were used to statistically test the hypothesis, the qualitative text analysis and digital ethnographic research components were used to inform and complement the quantitative analysis. More details on data collection and analysis methods can be found in Chapter 2.

Thesis Outline

This thesis comprises six chapters, which are based in part on five peer-reviewed journal and book chapter publications. Following the Introduction, Chapter 1 reviews the

existing literature on the socio-psychological drivers of radicalisation towards violence and will be published as a chapter in the *Routledge Handbook of Radicalisation and Countering Radicalisation* (Ebner & Whitehouse, 2023). It uses a historical lens to show how radicalisation frameworks have changed from linear one-dimensional models towards multi-dimensional ones, and how this has been mirrored in CT policies and deradicalisation approaches. It argues that there is not one pathway towards violent extremism, and that neither ideologies nor poverty or education levels alone offer sufficient explanation. Identity, however, has been a recurring and important factor that bridges the gap between grievance-led and narrative-led explanations.

Chapter 2 provides a detailed description of the materials and methods that were used for this project, situating the chosen analytical layers within the wider body of research by offering a discussion of existing studies that align with this project's methodological choice. Most notably, different definitions, methods, and innovations related to NLP analysis, qualitative text analysis, and digital ethnographic research are explored. The chapter also outlines the data collection and data management processes used for this thesis, as well as providing insights into the ethical standards that were applied. Finally, it explains the research limitations of the project based on the methodological framework and data availability.

In Chapter 3, the thesis presents the findings of the systematic analysis of 15 violent and non-violent manifestos. The first part of this chapter, which uses qualitative text analysis grounded in fusion theory to assess the manifesto content, was published in the peer-reviewed journal *Studies in Conflict & Terrorism* under the title "Is there a language of terrorists? A comparative manifesto analysis" (Ebner et al., 2022a). The

second part of this chapter, which illustrates the statistically significant narrative and linguistic patterns identified in the quantitative NLP analysis of manifesto content, has been submitted to another academic journal. Based on this mixed methods analysis, a new text-based violence risk assessment framework was designed.

Chapters 4 and 5 apply the new fusion-based coding framework to a range of far-right extremist groups along the violence spectrum, as well as to non-extremist control groups. Chapter 4 provides a comparative analysis of far-right extremist groups, providing both quantitative results from the NLP analysis and findings from the qualitative text analysis and digital ethnographic research in these virtual groups. It aims to test the new coding framework by comparing its estimated risk of violence for each group based on the linguistic markers grounded in fusion theory to the real-world links the assessed groups have to violent terrorist groups and activities. This chapter's contribution has been submitted to a peer-reviewed journal.

Chapter 5 assesses the risk of violence posed by ideologically fluid movements such as the QAnon conspiracy myth community. It uses the new coding scheme to analyse content from virtual QAnon groups and to assess the scale and nature of violence-predicting narratives within the movement. It quantitatively and qualitatively examines to what degree the messages in QAnon groups carry the trademarks of violent terrorist manifestos that are not found in non-violent texts. The narrative and language results for the QAnon groups community are then compared to content both from a non-violent group and a violent-terrorist control group. A peer-reviewed article based on the findings of this chapter has been accepted for publication in the journal

Perspectives on Terrorism under the title “The QAnon security threat: A linguistic fusion-based violence risk assessment framework” (Ebner et al., 2022b).

In Chapter 6, a comparison and discussion of the findings from all investigated groups is provided. This chapter outlines overarching patterns that were identified in the various sub-studies and examines their relationship to the fusion-plus-threat model. It aims to weave together the findings of all the previous chapters and discuss them within the context of previous fusion research. It also summarises the research findings before offering recommendations for future research and distilling policy implications. This final chapter also highlights future challenges related to the study of online pathways towards violence and discusses ethical considerations regarding the practical implementation of this study’s findings.

Scholarly Contribution

While this doctoral project conceptually and methodologically draws on various disciplines, it primarily seeks to make a scholarly contribution to the field of terrorism studies. This thesis brings a novel approach from social psychology to the assessment of extremist online groups and the study of violence indicators. It performs one of the first systematic theory-driven terrorist manifesto analyses and develops a violence risk index that addresses one of the key questions in terrorism research: what is the link between online discourse and offline violence? The following chapters will outline existing theoretical models in the field of terrorism studies, situate the socio-psychological approach used in this thesis, and outline the choice of methods inspired by digital anthropology and internet studies.

Chapter 1: Socio-Psychological Radicalisation Drivers

What are the distinctive cognitive and behavioural patterns associated with pathways towards violence? How do personal and group identities contribute to the process of radicalisation? What kinds of social environments and shared experiences may be conducive to violent extremism? This chapter addresses these questions in the broader context of psychological perspectives on radicalisation and terrorism. It will look across different disciplines – including terrorism studies and criminology, sociology and psychology as well as communication and internet studies – to identify relevant existing research on the socio-psychological drivers of radicalisation towards violence.

Psychological Explanations of Radicalisation

In the aftermath of 9/11, researchers and policy experts were divided over the question as to whether violent extremism was motivated primarily by ideologies or by unresolved grievances. These two possibilities are commonly referred to as "Kepel vs Roy", named after the two French scholars and former friends who came to disagree about the underlying causes of radicalisation (Nossiter, 2016). Gilles Kepel, a political scientist and Arabist who teaches at Sciences Po in Paris, championed the idea that ideology and collective narratives play a key role in the process of radicalisation towards violent extremism (Kepel, 2003). Oliver Roy, another leading French scholar and professor at the European University Institute in Florence, stressed the psychological state of suffering, as well as the desire for recognition, that lead individuals to embrace revolutionary youth movements that promise heroism and brotherhood (Roy, 2015).

While Kepel viewed ideology as the main driver of violent extremism, Roy saw ideological frameworks merely as outlets for personal struggles. In the context of jihadism, Kepel spoke of a "radicalisation of Islam" (Kepel, 2015), while Roy argued that we are witnessing an "Islamisation of radicalism" (Roy, 2017).

Although both Kepel and Roy focused their work predominantly on jihadist violence, their lines of argumentation have been applied to other forms of violent extremism and reflect a wider debate in radicalisation research. The first two decades of the 21st century were marked by a strong focus on radicalisation towards violent jihadist causes. Research has frequently highlighted the role of identity but tended to emphasise its ideological underpinnings (for example, Muslim identity politics and religious identification) as key elements of identity and driving forces in the radicalisation process (Choudhury, 2007). Importantly, this research pointed to underlying similarities in the radicalisation process across a diversity of groups (Bailey & Edwards, 2017; Lee & Knott, 2020), ranging from jihadist to far-right extremists (Ebner, 2017). This research brought to light various ways in which the narratives of violent extremists connected personal grievances to group-based struggles (Fielitz et al., 2018). In line with this, many researchers are increasingly viewing radicalisation as an identity-transformation process that incorporates both personally transformative and ideologically shared components (Whitehouse, 2018).

High-profile political engagement and media reporting on the topic of radicalisation have made it difficult to distinguish between the drivers and dynamics that are well-established in scientific research and those that are merely gaining traction in the public discourse but have little empirical basis (e.g. explanations based exclusively

on personality profiles or psychopathological considerations (Kruglanski & Fishman, 2009). Researchers at the Royal United Services Institute (RUSI) conducted an extensive review of post-2000 literature and academic articles about radicalisation to classify common hypotheses about pathways to extremism according to their scientific credentials (Allan et al., 2015). They found little or no support in existing studies for the arguments that masculinity and honour play significant roles in radicalisation. Likewise, the idea that low levels of literacy and the limited availability of information drive extremism is not empirically well supported. While the role of education, employment, poverty, and blocked political participation may play a role, the evidence is mixed. Likewise, efforts to establish distinctive terrorist and extremist profiles have proven to be inconclusive (Gill et al., 2014; Gruenewald et al., 2013; Neumann, 2017).

Researchers studying radicalisation have increasingly shifted their focus toward more evidence-based approaches (Ahearn et al., 2020; Gøtzsche-Astrup, 2018; Smith, et al., 2016). Based on a systematic review of the scientific literature on radicalisation towards violent extremism across different ideologies and geographies, Vergani et al. (2018) proposed the "3Ps of radicalisation" model, differentiating "push", "pull", and "personal" factors in the radicalisation process. These authors concluded that pull factors, which include cognitive drivers, social mechanisms, and group processes, as well as emotional and material incentives, were the most regularly emphasised in the literature. There is substantial credible evidence that identity is a key element in radicalisation (Allan et al., 2015). In particular, research has identified a wide range of factors associated with group psychology, especially the dynamic interplay between personal and group identities in motivating violent extremism.

Psychological Theories of Radicalisation

Social psychologists have developed a range of models that seek to explain radicalisation, using a combination of individual and social attitudinal and behavioural factors. Proposed psychological drivers include "frame alignment" (Wiktorowicz, 2005), "uncertainty–identity theory" (Hogg & Adelman, 2013), the "quest for significance model" (Kruglanski et al., 2018), the "staircase to radicalisation model" (Moghaddam, 2005), the "four-stage process" (Sageman, 2008), the "duplex theory of hate" (Sternberg, 2003), the "moral disengagement theory" (Bandura, 2004, p. 121), the "attitudes-behaviours corrective (ABC) model" (Khalil et al., 2019), the "two-pyramids model" (McCauley & Moskalenko, 2017), and the "fusion-plus-threat model" (Whitehouse, 2018).

Psychological approaches suggest that the perspectives of the two French theorists Kepel and Roy are not as incompatible as they and their followers have tended to assume. While Roy was right to point to the importance of shared suffering as a major source of motivation to engage in extreme forms of pro-group action, Kepel's collective narratives also play a role in motivating ideological extremism. Collective narratives can contribute to perceptions of shared suffering and thereby contribute to the deep psychological motivations that lead to violence. Based on extensive research into the British Islamist extremist organisation Al Muhajiroun, Wikotorowicz (2005) called the matching of individual factors and group factors "frame alignment", i.e. when an individual's "cognitive opening" coincides with the narrative offered by an extremist group. However, there is growing evidence that neither demography nor ideology are capable of motivating extreme self-sacrifice on their own. Research on radicalisation is

increasingly moving away from viewing radicalisation as a one-dimensional, linear process towards a more holistic, multi-level approach to framing extremism (Allan et al., 2015; McCauley & Moskalensko, 2008; Pisou & Ponesh, 2016; Ranstorp, 2010).

In 2005, the psychologist Fathali M. Moghaddham introduced his famous "staircase to radicalisation" model. According to this model (Moghaddam, 2005), most people inhabit the ground floor, which stands for perceptions of injustice, but some then move on to higher floors, from trying to find options to resolve the perceived state of unfair treatment (first floor) to the channelling of aggression towards an enemy figure or group (second floor). In the next steps, radicalised individuals would become morally engaged to justify hostility towards a demonised out-group (third floor) and increasingly endorse a terrorist organisation as their categorical thinking solidifies (fourth floor). Few people reach the uppermost floors, actually committing acts of terrorism and violence against out-groups (fifth floor). Lygre et al. (2011) found empirical evidence for Moghaddam's "staircase" model but further research is needed to clarify all the predictions of the model, including the key question of causality; in particular, what drives individuals to ascend the staircase that leads to violent extremism?

Another influential model, developed by Marc Sageman, conceptualises radicalisation as a four-stage process: "a sense of moral outrage, interpreted in a specific way, which resonates with one's personal experiences, and is channelled through group dynamics, both face-to-face and online" (Sageman, 2008, p. 223). While the model offers a useful framework to understand the connection between personal grievances and group narratives, it lacks an evidence-based explanation of the transitions between the four stages.

Social identity theory helps to explain how social identification and group membership can motivate and influence individual behaviour (Tajfel & Turner, 1979). Many studies have explored how strong group alignment (based on shared attributes such as class, party, gender, ethnicity, and religion) is linked to radicalisation (Beelmann, 2020; Harris et al., 2014; Smith et al., 2016). Strindberg (2020) argued that "group identity draws the battlelines" and that social identification processes are central to pathways towards extremism. Uncertainty–identity theory is based on the premise that uncertainty about one's self is a strong motivator for joining extremist groups that offer high levels of group entitativity and identification (Hogg & Adelman, 2013). There is substantial evidence in support of the causal links between identity uncertainty and group identification (Gøtzsche-Astrup, 2018).

It is widely recognised by social identity theorists and terrorism researchers that extreme attitudes and behaviours entail "us and them" thinking, which expresses favourable attitudes towards the in-group and negative attitudes towards the out-group (Harris et al., 2014). Berger (2017) adopted a framework based on social identity theory for his assessment of ideological shifts toward extremism and the construction of in-group/out-group dichotomies. Drawing on Ingram's (2016) linkage concept, Berger argues that extremists link out-groups to a threat and the in-group to a vulnerability, resulting in a crisis construct that may sit on a spectrum from mild to apocalyptic.

Kruglanski et al. (2018) introduced the so-called "quest for significance model", which combines three different factors to explain radicalisation: psychological need; ideological narrative; and social network (the 3Ns). The authors argue that an individual's need for personal significance or desire "to be someone" is the key factor

underlying radicalisation, which is then expressed through a narrative that promises a pathway to achieving significance through violent action in support of a collective cause and a network of people who similarly endorse the narrative (Kruglanski et al., 2018). A number of interview-based and observational studies have found evidence for the quest for significance approach (Neumann, 2015; Speckhard et al., 2017). However, both the causality and directionality of the triangular relationship have not been tested systematically.

The ABC model problematises the concept of radicalisation, emphasising that violent behaviour cannot be predicted on the basis merely of attitudes or ideology (Khalil et al., 2019). Like the ABC model, the "two-pyramids" model by McCauley and Moskalenko (2017) assumes two separate forms of radicalisation, namely radicalisation in opinion and radicalisation in action, with only the latter giving rise to terrorist activity.

While many of the models reviewed above provide useful frameworks and data for exploring the different psychological factors that have reappeared as patterns in radicalisation, they also leave many questions unanswered. Most importantly, they do not satisfactorily distinguish between psychological causes of violent extremism and mere correlates or consequences of the radicalisation process. This constitutes a major problem for the detection and prevention of criminal intent. Many individuals may exhibit beliefs and behaviours associated with radicalisation that are not in themselves illegal and may never lead to acts of violent extremism.

Is there one underlying factor that drives violent manifestations of radicalisation as distinct from merely extremist beliefs and attitudes? Is there some special

concatenation of features that, once combined, leads to acts of terrorism? Or are all the features currently associated with radicalisation mere epiphenomena that are caused by the same processes that drive violent extremists but also many harmless ideologists and bigots? If so, how can we grasp causation and what can we do about it? This thesis is based on the hypothesis that "identity fusion" is a key factor in causal pathways to violent extremism (Swann et al., 2009), including in online environments.

Online Radicalisation Models

In the past two decades, the internet and social media have added a new layer to the challenge of understanding the complex psychological and social processes that take place when individuals or groups radicalise. There is a rich and growing body of research on radicalisation towards violent extremism in the digital space, ranging from psychological experiments and sociological studies to linguistic analyses and narrative research in online extremist forums.

While the impact of the internet on radicalisation processes remains subject to controversial academic and political debates, there is now consensus that extremist recruiters and propagandists have been early adopters in exploiting the known bugs of new media ecosystems (Meleagrou-Hitchens & Kaderbhai, 2018). Over the past few years, the overwhelming majority of research projects on online radicalisation have focused on the propaganda, recruiting, and communication strategies used to lure individuals into jihadist and Islamist networks (Meleagrou-Hitchens & Kaderbhai, 2018). Most recently, the question of how algorithms, filter bubbles, and the prevalence of

disinformation produce online polarisation and fuel radicalisation has attracted much scholarly interest (Tucker et al., 2018).

In 1972, the psychologist Clark McCauley showed in an experiment that groups tend to adopt more extreme attitudes in the course of discussion. According to him, these so-called “group extremity shifts” strongly correlate with group opinion convergence (McCauley, 1972). More recent research has revealed that group dynamics that lead to the adoption of extremist views and behaviour can be accelerated through the infrastructural, political, cultural, and social implications of new online information and communication networks (Awan et al., 2011). For example, online anonymity has loosened social constraints and decreased social costs, leading to what the US cyber psychologist John Suler calls a “toxic disinhibition” effect (Suler, 2014). In 2010, Warner (2010) furthermore found that exposure to exclusively like-minded attitudes in so-called online filter bubbles may fuel polarisation and extremism.

Echo chambers are a highly contested topic in media and internet studies. The academic literature even diverges on the question of whether echo chambers exist or not (O'Hara & Stevens, 2015). The metaphor of the echo chamber has become a buzzword to describe networks of ideologically like-minded people, who remain isolated from information incongruent with their beliefs and/or interests and share information that confirms and reinforces their existing beliefs and/or interests (Jamieson & Cappella, 2010). Sunstein (2007) was among the first researchers to warn of online echo chambers and their effects on group polarisation.

In recent years, however, some researchers have cast doubt on the role of the internet and social media in creating politically fragmented communities. O'Hara and Stevens (2015) argued that the evidence in support of digital echo chambers is not strong or conclusive enough to justify regulation or intervention. Dubois and Blank (2018) of the Oxford Internet Institute (OII) even speak of "the myth of the echo chamber", showing that political interests and media consumption habits remain diverse among UK adult internet users, hence challenging the idea that social media leads to high segregation based on interests and opinions. These findings are supported by the work of Guess et al. (2018), who concluded that echo chamber effects are much more limited in scope and severity than popular discourse suggests, with the majority of US netizens continuing to keep diverse information diets. An analysis of the web-browsing histories of 50,000 US-based users came to similar conclusions about the exposure to politically diverse materials but showed that social media channels went hand in hand with an increased mean ideological distance between individuals (Flaxman et al., 2016).

While the prevalence and "information cocoon effects" in echo chambers may have been exaggerated, studies across different social media platforms and channels have demonstrated effects regarding the consequences of online networks for polarisation and radicalisation. For example, a Harvard study of the narratives and conspiracy theories shared by US and Italian Facebook users found quantitative evidence for the existence of hyper-polarised echo chambers (Quattrociocchi et al., 2016). Likewise, Jonathan Bright of the OII illustrated in a large-scale study of 90 different political parties on Twitter across 23 different countries that, with mounting

ideological difference between parties, the likelihood of interaction between their networks decreased. This correlation was especially significant among the extreme fringes (Bright, 2017). The majority of studies have focused on media and information consumption patterns of general internet users rather than attitude and behaviour trends within specific online networks and sub-cultures. Most research into online echo chambers has furthermore taken a categorical, one-dimensional approach to defining and measuring echo chambers. As a result, the multiple forms that echo chambers can take, both in their nature and their intensity, have not been captured well in existing measures.

Most scholars suggest that the exploitation and operationalisation of grievances by extremist propagandists and recruiters plays a significant role in online radicalisation campaigns (Kundnani, 2012a; Neumann, 2013). A recent study by the ISD found that supporters of the white nationalist "Unite the Right" rally of August 2017 shared a range of grievances related to race, the political left, freedom of speech, Southern heritage, and "the establishment" in the run up to the protests in Charlottesville (Davey & Ebner, 2017). Research into cumulative extremism and reciprocal radicalisation dynamics has furthermore shown that individuals in extreme-right forums tend to share their negative experiences related to minority communities and discuss witness accounts of traumatising events, such as jihadist attacks and sexual abuse cases. Similarly, members of Islamist extremist networks tend to reflect on their experiences during the war on terror in Muslim-majority countries, "Western" war atrocities against Muslims, anti-Muslim hate crimes, and other experiences of discrimination in European countries (Crawford et al., 2018; Fielitz et al., 2018).

Despite a wide range of studies aiming to understand how individuals and groups adopt extreme attitudes and behaviours in the digital space, questions concerning whether and how group identification and identity fusion processes take place in extremist forums online remains an enigma. In particular, systematic studies of the interaction between virtual identity dynamics and pathways to violence are still largely absent from the existing body of research. This project therefore seeks to assess the relationship between identity transformations and radicalisation towards violence in online groups.

Identity Fusion: An Emerging Psychological Explanation of Violent Extremism

As explained earlier, social identity theory fruitfully distinguishes two kinds of identity: the personal self (comprising traits that make us unique as individuals); and the social group (comprising traits that mark off a collective of some kind, whether it is a small relational one, such as a family or a football team, or a large categorical one, such as a nation, ethnic group, or world religion). Social identity theory is centred around group identification, which commonly presents personal and group identities as quite separate and unconnected. However, an alternative form of group alignment is presented by identity fusion, which entails a porous boundary between personal and groups identities (Swann et al., 2012). This section draws on a growing body of evidence to argue that these two forms of group cohesion tend to lead to different outcomes in extremism: cognitive versus violent extremism.

Identification is derived from sharing identity markers, such as group beliefs and practices, whereas identity fusion taps into personal agency in a more visceral way

(Whitehouse, 2018). For example, traumatising or deeply transformative events shared with other members of an in-group can help to create the synergistic relationship between self and group that is the hallmark of identity fusion (Whitehouse et al., 2017). Since fused individuals tend to view members of their in-group as kin-like (e.g. a brotherhood), metaphors of shared blood applied to other group members are among the most prominent linguistic indicators of fusion (Whitehouse & Lanman, 2014).

Identification with a group is depersonalising in the sense that making the group salient will cause the personal identity to become less accessible, and vice versa (Tajfel & Turner, 1997). For example, when you think about your identity as a Muslim (e.g. the beliefs and practices associated with that group), your unique personal identity will become less relevant. And the same applies the other way around: if you think about your unique self as an individual, your sense of yourself as just another member of a big religious organisation will be less relevant. With identify fusion, however, it is a very different story.

When people fuse with a group, their personal and group identities are activated synergistically, rather than hydraulically (Swann & Buhrmester, 2015). If you make the person's identity as a Muslim salient, it activates their sense of personal agency at the same time. Conversely, if you make their personal identity salient, that taps directly into their conception of what it means to be a Muslim. This fusion of personal and group identities motivates extremely strong forms of pro-group action. Most of the time, identity fusion is a positive thing. For example, many people are fused with their families, and are therefore willing to make great sacrifices for their relatives, whether by giving them help financially or, in more extreme cases, taking care of them during

prolonged illnesses or donating organs when transplants are required. But there is also a dark side to fusion. Because the self and the group are so tightly bound together, when the group is threatened, fused individuals take it personally. As a result, they will stop at nothing to defend the group, displaying far greater willingness to fight and die for each other even than the most highly identified individuals.

While strong forms of identification with a group can motivate commitment to extremist ideologies, this form of group alignment does not appear to be sufficient to motivate extreme forms of self-sacrifice (Swann & Buhrmester, 2015; Swann et al., 2012; Whitehouse, 2018). Self-preservation is a strong motivating force in humans and other animals, overriding coalitional commitments when they imperil survival. In contrast, since fused individuals integrate the group into their sense of self, they are more likely to throw themselves on a grenade to save their fellows (Gómez et al., 2011; Swann et al., 2010). Evidence for this comes from a variety of studies, including fieldwork, surveys, and interviews with armed groups that regularly risk life and limb to protect their fellow group members. For example, fusion has been linked to extreme pro-group violence in Libyan revolutionary battalions (Whitehouse et al., 2014), Cameroonian herders and farmers (Buhrmester et al., 2020), Indonesian religious fundamentalists (Kavanagh et al., 2020), British and Brazilian football hooligans (Newson, 2019), and captured fighters from Islamic State and Kurds on the battle frontlines (Atran, 2016).

The fusion-plus-threat model integrates previous inter-disciplinary evidence about the drivers and consequences of fusion and identification into a broader theory of "ritual modes" in which differing memory systems give rise to contrasting forms of

group alignment (Whitehouse, 2018). The theory proposes that an emotional event stored in episodic memory, and usually accompanied by the individual's reflection on its meaning, can contribute to personal identity transformation. When such a transformative experience is perceived to be shared with a group, this helps to create a feeling of shared group essence that can produce fusion. A second pathway to fusion is based on phenotypic matching, hence perceived biological kinship (Whitehouse et al., 2017). These pathways to fusion and self-sacrifice are depicted in Figure 3.

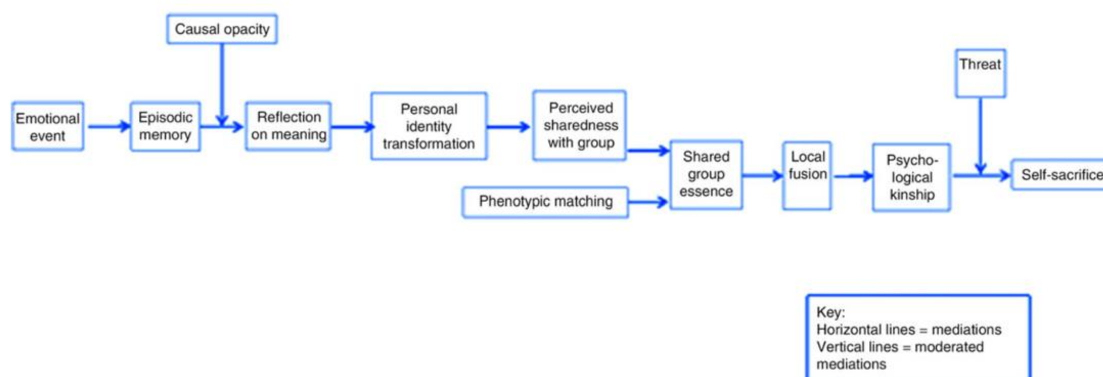


Figure 3: Pathways to Fusion and Self-Sacrifice (Whitehouse, 2018)

Both pathways (the sharing of personally transformative experiences and the sense of biological kinship) lead to a form of psychological kinship and fusion with other group members. When faced with an external threat to the in-group, identity fusion can significantly heighten an individual's willingness to fight and die for the group. Fusion therefore offers an explanation for acts of self-sacrifice, such as suicide terrorism and martyrdom operations, as well as other forms of extremely risky pro-group behaviours (Whitehouse, 2018).

Based on the open commentary in response to Whitehouse's fusion-plus-threat model, several variables that might mediate or moderate pathways from fusion to violence have been identified (for a fuller account of this framework, see Whitehouse, 2021). As outlined in the Introduction, the following factors have been frequently cited in the literature as contributing to an escalation towards violence: perceived out-group entitativity (Choi et al., 2018; Lickel et al., 2006); perceived out-group threat (Whitehouse, 2018); and violence-condoning norms (Louis et al., 2018).

Measuring Forms of Group Cohesion

Anthropologists and social psychologists have long tried to measure the consequences of group cohesion for individual and collective behaviour, as well as for inter- and intra-group dynamics (Cartwright & Zander, 1968; Tajfel & Turner, 1997). To test different sets of hypotheses about this relationship, researchers have developed a variety of measures to assess the different forms of social alignment. While early measures of identification from the 1980s and 1990s tended to be unidimensional (Doosje et al., 1995), the turn of the millennium saw a range of multi-dimensional frameworks to measure different components of identification (Ashmore et al., 2004; Cameron, 2004; Leach et al., 2008). For example, Leach et al. (2008) identified five specific components of in-group identification from the existing literature and grouped them into two overarching dimensions to build their model of in-group identification: (1) self-definition (individual self-stereotyping and in-group homogeneity); and (2) self-investment (solidarity, satisfaction, and centrality).

Drawing on Mael and Ashforth's (1992) reformulated model of organisational identification, Mael and Tetrick (1992) first measured identification with a psychological group (IDPG) based on a pilot study of 161 undergraduates, drawing a clear distinction between organisational identification and organisational commitment. Regarding these two factors, their scale measures the endorsement of claims such as "This organization's successes are my successes" (i.e. shared experiences) and "If a story in the media criticized the organisation, I would feel embarrassed" (i.e. shared characteristics). More than 10 years later, Postmes et al. (2013) developed a single-item measure of social identification, which captures the essence of group identification with a seven-point scale for the statement "I identify with my group (or category)". These authors tested the model across a wide range of social groups, finding that its convergent and divergent validity as well as its test-retest reliability were high. Even though short scales have traditionally been regarded as methodologically less rigorous, they tend to have better conceptual clarity, work well in homogenous, clearly defined contexts, and can prove especially useful when respondents have limited time to answer questions.

As mentioned in the previous section, social psychologists initially explained changing identity notions in intergroup conflict as a process of deindividuation (Reicher et al., 1995). However, more recent research has highlighted the concept of identity fusion (Swann et al., 2009, 2012; Whitehouse & Lanman, 2014), which is theoretically grounded in social identity theory (Tajfel & Turner, 1979). This led to the development of metrics to measure identity fusion. Swann et al. (2009) introduced a pictorial measure that visualises the degree of fusion of personal and social identities. These authors used a measure that was originally developed to determine attachment in close relationships

(Aron et al., 1992) and then modified to measure group identification (Coats et al., 2000). Stressing the shortcomings of this single-item pictorial measure, Gómez et al. (2011) designed a seven-item verbal scale that captures the two key features of fusion: (1) perception of connectedness with the group; and (2) reciprocal strength. Most recently, researchers have developed the so-called dynamic identity fusion index (DIFI), which uses the single pictorial item developed by Swann et al. (2009) but provides a higher resolution by using a continuous scale that assesses both distance and overlap between two circles that represent “the self” of the respondent and “the group” (Jiménez et al., 2016).

To conclude, previous research suggests that two forms of group alignment, identification and fusion, represent separate strands of radicalisation, one rooted in highly shared ideology and the other stemming from highly shared self-defining experiences, which can give rise to violent pro-group behaviour (Gómez et al., 2011; Swann et al., 2009; Whitehouse, 2018). This helps to explain why extreme pro-group behaviour can be developed without adherence to extremist ideologies and, vice versa, why the belief in extremist ideologies is not necessarily tied to the willingness to commit acts of violence. This research project seeks to test this hypothesis in the virtual world, assessing in particular their validity for new forms of groups in online settings.

The findings regarding identity fusion via the sharing of dysphoric experiences as a driver for extreme pro-group violence are compatible with narrative-based research into extremist movements. Perceived collective victimhood (Lynch & Joyce, 2018) and victimisation narratives (Pemberton & Aarten, 2018) are considered key elements in the study of radicalisation. For example, Pemberton and Aarten (2018) examined the

relationship between victimhood, narrative, and identity, arguing that the radicalisation path is shaped by a narrative understanding of identity that evolves around victimisation experiences. Likewise, Crossley (2020), whose research focuses on the intersection of narrative psychology and identity, concluded that traumatising experiences constitute narrative ruptures that can shatter an individual's sense of self-continuity.

As outlined in the Introduction, the key aim of this doctoral thesis is to make an academic contribution to the field of terrorism studies. The research project is an attempt to bring together the different strands of research on radicalisation towards violence by combining findings from the disciplines of psychology, sociology, anthropology, and communication studies. Although it focuses on the role of identity dynamics in the radicalisation process, the underlying assumption is that identity provides the bridge between grievance- and narrative-focused research. Different pathways towards extremist attitudes and behaviours are all characterised by identity transformation processes that reshape an individual's relationship to an online group or loose virtual community. This identity transformation takes place either in the form of group identification, identity fusion, or a combination of the two, determining the kind of extremism it fosters (see Figure 1).

Chapter 2: Thesis Methodology

This chapter outlines the methods, tools, and materials used in this research project. It also situates the methods selected for this thesis within the wide array of existing approaches to data collection and analysis, as well as discussing ethical considerations that were relevant to this project.

Over the past decade, social media has become a central force of change for our societies, both on an institutional and a human level. As with previous disruptive innovations in the communication and information space, such as the printing press, the telephone, or the personal computer, its impact on the identities of individuals, organisations, and nations may be even more powerful and multifaceted than we can currently imagine. The footprint that social media and its new opportunities leaves on our societies extends from changing people's media- and information-consumption habits and giving rise to new subcultures and protest movements to inspiring hatestorms, terrorist attacks, and election interference campaigns (Bartlett, 2018; Fielitz & Marcks, 2019; Marwick & Lewis, 2017).

As a result, governments, researchers, and private sector companies have sought to better understand the emerging virtual worlds and invested in developing online analysis capacities. A range of tools have been created to map the networks of online populations, analyse their conversations, and examine their ideological leanings. Everything, from a certain sub-population's opinion on Boris Johnson to their preferred brand for sunglasses, can be tracked and assessed. These tools, which allow for the mapping of attitudes and behaviours of individual users and entire online communities,

have enabled private companies and politicians alike to develop new forms of micro-targeting. Using the analysis of netizens' habits, desires, and language patterns, it is possible to find new modes of persuasion to buy certain products or vote in a certain way (Breese, 2015).

Some researchers have warned that political parties across Europe and the US have used tactics of disinformation, deception, and hyper-targeted ads in their campaigns to reach specific social media populations and manipulate them for the sake of election gains (Applebaum et al., 2017; Colliver et al., 2018). Other scholars, however, have argued that the influence that micro-targeting and online campaigning have on voting behaviours is overstated (Hersh & Schaffner, 2013; Kalla & Broockman, 2018; Krotzek, 2019). For example, evidence from 49 field experiments suggests that the persuasive effects of campaigning only apply when the candidates hold unpopular positions, when there is strongly investment in identifying persuadable voters, and when voters are contacted long before election day (Kalla & Broockman, 2018). In 2010 and 2011, Eitan Hersh and Brian Schaffner conducted survey experiments that presented the respondents with a fictional election candidate and mock campaign mailers. The results of the study indicated that "mistargeted voters" penalise candidates enough to offset the positive effects of targeting (Hersh & Schaffner, 2013).

The present project made use of digital data-gathering and -management techniques that can also be exploited for political purposes by studying the desires and fears revealed through the language used online. However, it used these approaches within the scope of the research questions relevant to this thesis. It is argued that the benefits of understanding what drives individuals towards extreme forms of violence

outweigh the costs of inaction. Data-driven research into the drivers and characteristics that dominate radicalisation processes in online spaces is an essential step for preventing future terrorist attacks and hate crimes. More details about the ethics approach to the management, analysis, and publication of the data collected in this project can be found in this chapter's sub-section "Ethical Considerations".

Data Mining

The data collected for this thesis can be grouped into two types: (1) a collection of manifestos by violent and non-violent authors, used in Chapter 3 to distil narrative and language patterns in the writings of terrorists and to design a fusion-based violence risk assessment framework; and (2) online groups across the violence spectrum that were scraped for the analysis performed in Chapters 4 and 5 to test the newly created narrative and language framework. The manifestos did not require any data-mining activities, as they were collected as PDF downloads over the course of the last decade and converted into text documents (TXT) to import them into R Studio for the NLP analysis. To source the data needed for the online group analysis, a professional web scraping service was used. The limitations of this approach will be discussed in Chapter 6.

The aim was to capture a wide range of public online spaces, including forums, chat rooms, and social media groups. In order to ensure an ideologically agnostic approach, virtual groups from across the violence and ideological spectrum are represented in the sample. To allow for sufficiently powered hypothesis testing using NLP analysis, all selected groups needed to contain a minimum of 1,000 posts that could

be scraped and used for the analysis. Specifically, data were collected from four types of online platforms: (1) the Iron March website, which was a violence-endorsing white supremacist discussion forum that was closed down due to its links to terrorist organisations such as Atomwaffen Division and National Action; (2) the online gaming chat application Discord, which has hosted a variety of ideologically extreme alt-right and conspiracy theory groups; (3) the moderate religious discussion platform Third Hour, which is marked by a shared moderate ideology but is not endorsing violence; and (4) the football fans' forum FootballForums.net, which is a popular non-ideological hobby community that hosts mainly non-violent football fans but may include a small number of hooligans who have resorted to violence in the past. Table 1 provides an overview of the scraped groups that will be analysed in Chapters 4 and 5.

Table 1: Overview of Scraped Online Groups

Group Name	**Group Explanation**	**Links to Violent Extremism**	**Cognitive Extremism Spectrum**	**Number of Scraped Messages**
Iron March	Iron March was a public web forum used by members of the neo-Nazi terrorist groups Atomwaffen Division and National Action, as well as the Scandinavian far-right extremist Nordic Resistance Movement and the Australian White supremacist Antipodean Resistance. The website was originally created in 2011 by Russian-based nationalist Alisher Mukhitdinov, who went by the name Slavros. Its users have been linked to several terror plots and violent extremist incidents. Iron March closed in 2017 and was	Mainly violent	Ideologically extreme	1,164

	then turned into a searchable database by the anonymous Twitter user @jewishworker in 2019 (Singer-Emery & Bray, 2020; Upchurch, 2021).			
Charlottesville 2.0	Charlottesville 2.0 was one of the key Discord servers used by the organisers of the white supremacist "Unite the Right" rally in Charlottesville, Virginia, on 11–12 August 2017. The protest brought together members of the alt-right, neo-Nazi movements, right-wing militias, and neo-Confederates. The rally escalated into several instances of violence, including the deadly car attack by James Alex Fields, which killed counter-protester Heather Heyer and injured 35 others. The server was active during the timeframe June to August 2017 before it was taken down by Discord. Its content was first leaked by Unicorn Riot (Schiano, 2017a).	Partly violent	Ideologically extreme	35,607
The Anticom Server	The Anticom Server was a White supremacist Discord server of the militant group Anti-Communist Action. Anticom had overlapping membership with the Atomwaffen Division and shared improvised explosives and combat manuals. The Anticom Discord channel was also used by participants of the "Unite the Right" rally in 2017. The server was active during the timeframe February to September 2017. Its content was first leaked by Unicorn Riot (Schiano, 2017a).	Partly violent	Ideologically extreme	188,302

The Right Server	The Right Server is a far-right extremist channel on Discord that was founded by former administrators of the extremist pro-Trump Centipede Central chat room. It was active during the timeframe 2017–2021. After The Right Server was deleted by Discord in 2018, a replacement channel continued under the new name The Right Goys, which is Yiddish for non-Jews (Liao, 2018). The messages of the original server were first leaked by Unicorn Riot (Alexander, 2018).	Partly violent	Ideologically extreme	678,246
The Great Awakening Community (QAnon Chat)	QAnon is a globally active conspiracy theorist network. The Discord server "Great Awakening Community (QAnon Chat)" was one of the main communication hubs for QAnon during the timeframe 2018–2019. The channel was first leaked by Unicorn Riot (Alexander, 2018).	Partly violent	Ideologically extreme	200,000
Third Hour	Third Hour is a popular discussion forum for Mormons. The so-called "Forum for Latter-Day Saints" is used as a place for Mormons to discuss topics such as family, marriage and parenting, missionary work, general welfare, and conversion stories, but also includes gospel boards and advice boards. There are no known links of forum members to violence or extreme interpretations of the Latter Day Saint (LDS) movement.	Non-violent	Ideologically moderate	161,977
FootballForums.net	FootballForums.net is a popular UK discussion forum for football fans. Its users	Mainly non-violent	Non-ideological	304,910

	discuss British and international football, including the UEFA Champions League and Europa League, the Premier League, and non-league football and domestic cups. They also have boards for off-topic discussions and more personal conversations. The forum is a mainstream football hobby forum with users who support a range of different football clubs and therefore not expected to include many radical football hooligans.			

As this project focuses on specific fringe forums, messaging boards, and chat applications rather than mainstream platforms such as Facebook and Twitter, scraping mechanisms needed to be adapted to the architecture of each analysed platform. Data-mining difficulties related to the selected platforms vary depending on the API access, as well as the lifetime of individual threads, posts, or forum entries until they are removed. Web-crawling and -scraping tools for identifying, downloading, and indexing relevant data minimise the manual work involved in the process both of the data scraping and the raw text analysis. They often use Java- or Python-based scripts, both of which allow for automated content data mining. Java is a statically typed language that has powerful networking capabilities and can be a highly efficient tool for systematic web browsing and indexing. Python, on the other hand, is a general-purpose language that integrates well with other tools and has better text-processing capabilities. However, Python is dynamically typed and can be slower than Java-based web crawlers.

Based on so-called "seed websites" (websites that are used as starting points), web-crawler programmes search the web, which currently consists of 1.5 billion websites, for related hyperlinks to extract and index. Ready-to-use web-scraping applications, such as Apache Nutch, StormCrawler, and Scrapestorm, allow for the automated, intelligent downloading of data in multiple formats, including Excel, CSV, TXT, HTML, Database, and WordPress. However, these applications' limitations include the fact that they scrape content in a standardised form that might not allow researchers to tailor the extracted data to their specific research queries. For this reason, many researchers prefer to write their own coding scripts. Over the last few years, social science researchers have increasingly borrowed methods from computer sciences to program their own tools tailored to their specific research projects and the platforms they want to study. Due to the widely varying architecture of the websites and forums that host the groups selected for this project, this project used a professional web-scraping service to collect the required data via tailored scripts for each of the above-mentioned platforms.

Data Storage

To ensure all mined data were adequately stored and managed, professional data preservation and archiving systems were used. All electronic field notes and mined data, including visual materials and screenshot evidence, were stored securely on an encrypted laptop for the duration of the analysis. Collected text-based content from websites was downloaded into Excel sheets, as they can be integrated with NLP analysis tools using the CSV file format. Online websites that might be removed due to platform policies were archived using archive.is, which allows users to preserve the content of

any given webpage by simply entering the URL into the system. Any data transfers were made via an encrypted Google Drive system that requires a username and password. All collected data were treated confidentially and held in accordance with the General Data Protection Regulation (GDPR), the Data Protection Act 2018, and research codes of conduct. Any identifiable data were immediately deleted, unless posts from a user fell under the Prevent Duty, in which case the data and relevant details were passed on to the responsible authorities. For at least three years after final publication, all data are stored on two separately kept encrypted external drives.

Natural Language Processing (NLP)

Originating in different disciplines, various strands of digital research have been used for the quantitative analysis of virtual spaces, including NLP and network analysis. NLP was a central methodological string of this research project to quantitatively assess the collected online content. NLP is an umbrella term for a variety of computational methods to automate the analysis of language. The following section provides an interdisciplinary and cross-sector overview of methods to analyse online networks, content, and behaviour. A few selected examples of research studies using different methods in different disciplines are mentioned to illustrate their wide range of applications. However, these example studies only represent a small sample and are not meant to be a complete guide to research projects that used digital methods.

The most widely used programming languages for NLP techniques are R and Python. While both R and Python lend themselves to sophisticated word and syntax analysis, as well as visualisations of the findings, Python has the additional advantage of

allowing for automated content scraping. NLP-based research has its origins in the 1950s and was initially focused almost exclusively on word count and syntax analysis. With the development of machine-learning capabilities, NLP is now also used to conduct complex topic modelling, as well as sentiment analysis (Nielson, 2011; Yi et al., 2003) and option-mining exercises (Cambria & White, 2014). NLP has also been extended to cover different languages, including Arabic (Habash, 2010) and Chinese (Huang & Chen, 1996).

Two main types of NLP analysis can be distinguished. The first type is semantic analysis, which aims to understand the meaning in language. The second type is syntactic analysis, which studies the structure or grammar of a sentence (Reddy & Hanumanthappa, 2014). A number of studies have focused on the use of NLP to detect online extremist and terrorist content using semantic or syntactical analysis (Mussiraliyeva et al., 2020; Torregrosa et al., 2022). A recent survey of research articles applying NLP approaches (including topic modelling, sentiment classification, and semantic analysis) to study extremism and related concepts (including terrorism, hate speech, supremacism, and fundamentalism) detected a total of 70 articles that has been published by 2022 (Torregrosa et al., 2022). In this thesis, an NLP-based word-count analysis using R is combined with manually supported semantic analysis that examines the meaning of the relevant linguistic markers within the context of the wider dataset.

Beyond the above-mentioned academic research strands, it is worth looking at methods and tools developed by private sector companies, such as marketing agencies and strategy consultancies, as well as security forces, intelligence services, and political campaigners, for the purpose of analysing the attitudes and behaviour patterns of online populations. Of particular relevance for this project are the social listening and social

media analysis tools that have been used by think tanks to study the trends in the attitudes, conversations, and behaviour patterns of online populations. For example, the Boston-based social intelligence provider Crimson Hexagon, which has now merged with Brandwatch, produces content analysis and visualisations of dominating narratives on social media (Brandwatch, 2018). The tool allows researchers and advisors to enter one or more keyword(s) or hashtag(s) to track how related content travels and spreads across the Twittersphere and forums such as 4chan and 8chan. Combining NLP and geolocation data for each search query, the tool provides an analysis of how the volume of posts using the relevant keyword or hashtag has developed over time and across different geographies, which other words and narratives they have been most commonly associated with, and which social media users have been most influential in spreading them.

Many of the social media analysis tools were originally built for private sector applications, as they provide market analysts with valuable information on customers' interests, habits, and desires, which is then used to inform business and investment decisions. However, they have increasingly been repurposed by think tanks and academic researchers to map political conversations or hateful speech across different geographies. While marketers and market researchers use these tools to find answers to questions such as "What do consumers think of a new product?" or "Did an advertisement reach and engage the intended audience?", academic and think-tank researchers employ them to address questions such as "How do individuals express opinions about a social or political topic?" or "How does the public respond to an event?" (Breese, 2015).

Compared to other social media analysis tools, Brandwatch offers a major advantage to researchers who want to conduct long-term studies: with one trillion saved social media posts, it holds one of the most comprehensive databases of historic data from social media platforms and forums since 2012 (Brown, 2017). Examples of research projects that have made use of Crimson Hexagon range from studies on Twitter discussions about nanotechnology (Runge et al., 2013) and mental health (McClellan et al., 2017) to anti-Americanism and anti-interventionism in the Arabic Twittersphere (Jamal et al., 2015). This tool has been used by the ISD to examine international extreme-right social media campaigns (Davey & Ebner, 2017), to track disinformation and manipulation attempts in the run up to elections (Avaaz & ISD, 2019; Colliver et al., 2018; ISD, 2019a), and to trace back the evolution of conspiracy theories such as the "great replacement" theory, which inspired the latest wave of extreme-right terrorist attacks in New Zealand, the US, and Germany (Davey & Ebner, 2019). However, a major disadvantage of the tool is that it predominantly captures content from large tech platforms but does not cover smaller alt-tech platforms and encrypted messaging apps.

While social listening and user-friendly social media analysis tools were not used as part of the core methodology of this project, the above-mentioned previous studies that were based on Brandwatch analysis are helpful in situating the groups selected for analysis within the wider online ecosystem. Potential follow-up studies of this thesis may integrate the linguistic analysis framework developed here with artificial intelligence (AI)-based tools. One of the leading AI-driven text analysis technologies is the Java-based tool Method52, which was developed jointly by the Center for the Analysis of Social Media (CASM) at the London-based think tank Demos. Method52 combines machine-

learning mechanisms with NLP capabilities and enables the quantitative assessment of the specific linguistic markers and language patterns of any scraped materials, including raw or unstructured texts that are scraped from any platform (Sage Research Methods, 2016). The AI-based algorithms can be trained for specific research projects to produce sophisticated analyses of large sets of data. Demos itself used the in-house technology, for example, to measure Islamophobia in reaction to terrorist attacks and the Brexit vote in 2016 (Miller, 2016), to assess the scale of social media misogyny (CASM, 2016b), and to monitor the online discourse ahead of the 2016 Irish general election (CASM, 2016a). Other Method52-based research includes a study of Islamist and tribalist online messaging during the 2017 elections in Kenya (Amanullah & Harrasy, 2017) and a report on the phenomenon of reciprocal radicalisation between online Islamist and far-right groups, whereby researchers trained the AI tool to identify different forms of inter-group hatred, in-group victimisation, and out-group demonisation narratives (Fielitz et al., 2018).

Digital Ethnographic Research

The quantitative research parts of this doctoral project were informed by, and complemented with, digital ethnographic research. Notably, digital ethnography and qualitative text analysis were used to gain a deeper understanding of the nature and context of the analysed content. The digital ethnographic research string of this thesis was inspired by approaches from several disciplines. Many researchers, from fields such as digital sociology, media, and internet studies, have developed models to assess the impact of the "digital age" on individuals and collectives. Deborah Lupton named four ways in which the discipline of sociology could be reshaped by new technologies: (1)

using digital tools for networking and communication among researchers and their informants; (2) researching how people's use of digital technologies influences their identities and social relations; (3) using digital tools for quantitative or qualitative data analysis; and (4) conducting reflexive analysis of new technologies informed by social and cultural theory (Lupton, 2014, pp. 15–16). Research conducted for this project comprised elements of both (2) and (3). Researching the impact of people's use of digital spaces and the development of virtual sub-cultures on their identities was a central component of this research project, while digital tools were used to conduct systematic analysis of online content.

In recent years, including during this doctoral project, the field of digital ethnography has been marked by rapid developments. Emerging popular social media platforms (e.g. Tik Tok and BeReal, as well as alt-tech platforms such as Odysee, Truth Social, Parler, and Bitchute) and new approaches to internet governance (e.g. the Digital Services Act and the German Network Enforcement Act) have changed the algorithmic infrastructure, communication modes, and policy landscapes that need to be considered in digital ethnographic work. The following sub-section will outline the historical evolution of the field of digital ethnography from the dawn of the internet to today's metaverse developments, as well as tying existing approaches to the methods used in this thesis.

Digital ethnography is a fairly new discipline. However, Escobar et al. (1994, p. 211) suggested as early as in the 1990s that "the cultural constructions and reconstructions on which the new technologies are based and which they in turn help to shape" should be studied as a new domain within anthropology. He called this new

subfield the “anthropology of cyberculture”. The emerging cyber spaces and communities, however, remained significantly understudied for the following decades (Coleman, 2010 , p. 488), and digital ethnographic works remained scarce (Wilson & Peterson, 2002, p. 450). Subsequently, Horst and Miller (2012) made a case for the introduction of digital anthropology as a sub-discipline in its own right.

While most scholars recognised the importance of the evolutions that new technology would bring to the field of anthropology, many early scholars merely focused on visions of how digital innovation could enhance the tools available to anthropologists to collect, record, and publish data. They also saw IT as an opportunity to help bridge the geographic divide between home and field sites and conduct interactive off-site fieldwork (Rheingold, 1993; Stone, 1991). Alan Macfarlane and Sarah Harrison created one of the first anthropological multi-media archives, the Thak Archive, which includes ethnographic records collected since 1969 from the Gurung village of Thak in Nepal (University of Cambridge, 2020). In 2000, the Digital Himalaya project was launched by Alan Macfarlane and Mark Ruin to preserve and publish old digital materials from the Himalayas, such as photos, video recordings, and journals (Macfarlane & Turin, 2020).

Gustaaf Houtman and David Zeitlyn were among the first anthropologists who noted that digital innovation is bound to impact the discipline on several levels, namely data, methodology, and subject matter (Houtman & Zeitlyn, 1996). Beyond their assessment that new media is revolutionising the means for acquiring, recording, transmitting, and publishing data, they also asserted that it would reshape the identities of communities. For instance, they argued that the internet had already led to shifts in the self-perceptions of minority groups in the Middle East. These two researchers

concluded that humans think of and perceive IT as an extension of themselves and their realities. Before the turn of the millennium, Turkle (1995) had already predicted that virtual worlds would allow individuals to adopt new forms of multiple identities, while Stone (1991) provided examples of “virtual cultures”, stressing the mutually dependent relationship between technology and culture.

The early 2000s then saw a shift in the focus of digital anthropology. Anthropologists increasingly started to understand digital media as an integral part of culture instead of viewing it merely a context for culture (Wilson & Peterson, 2002). Crystal (2001) even suggested that the internet and new technologies might cause a “linguistic revolution”, giving rise to entirely new languages. Today, entire think tanks and research institutions are dedicated to studying online subcultures, encompassing their communication and linguistic specificities. For example, the Data and Society Research Institute (2019) used digital ethnographic research to create a better understanding of the social, cultural, and ethical dimensions of online environments. Some researchers, however, continued to perceive cyber cultures as too ephemeral and short-lived to be studied in isolation from offline contexts (Wilson & Peterson, 2002, p. 456).

With the emergence of the first social media platforms and online subcultures, such as the 4chan community of trolls, gamers, web-cam girls, and the Anonymous collective of hackers, researchers started devoting attention to studying networked communities (Coleman, 2010; Postill, 2011), virtual worlds, and fully immersive online environments (Boellstorff, 2008; Nardi, 2010). Coleman (2010, p. 490), for example, wrote that we are seeing “cultivated new modes of communication and selfhood;

reorganized social perceptions and forms of self-awareness; and established collective interests, institutions, and life projects". Horst and Miller (2006) additionally found that digital technologies such as mobile phones can extend and intensify existing kin networks.

Participative virtual communities can also take the shape of fully immersive online environments. Such new communities raise questions about the norms and social relations, including forms of governance and economic exchange, within such communities (Schroeder, 2011). Ralph Schroeder examined Active Worlds and Second Life as examples of "multi-user virtual environments" (MUVEs), describing new forms of interactions among large online populations in worlds for gaming and socialising as "being there together" (Schroeder, 2011).

To keep up with these newly emerging virtual worlds, digital ethnography has become more than an application of ethnographic practices to online environments (Pink et al., 2016, pp. 2–4). Methods might go beyond traditional ways of both communication and documentation used for anthropological studies. Participant observation can be conducted by reading posts or watching videos and recorded in the formats of photography, video, or blogging (Pink et al., 2016, pp. 2–4) but it can also involve setting up an online identity that is detached from the researcher's real identity to gain access to exclusive virtual communities (Bluteau, 2019). This project will make use of avatar accounts to establish an online identity and maintain a presence in public extremist forums and groups. However, due to the nature of the subject and its related security threats and safeguarding duties, participant observation will be limited to observation only, as any form of active participation or interaction could be ethically

unsound (see this chapter's sub-section "Ethical Considerations" below).

One of the biggest debates in the field of digital ethnography concerns the question of whether online communities and cultures are real or imagined (Bordieu & Colemen, 1991; Rheingold, 1993; Wilson & Peterson, 2002). The argument made by Benedict Anderson (1983, p. 49) in *Imagined Communities* concerning nations being social constructions, namely that "because the members of even the smallest nation will never know most of their fellow-members, meet them, or even hear of them, yet in the minds of each lives the image of their communion", also applies to cyber cultures, where most users will never meet in real life. In the early days of the internet, anthropologists regarded the virtual world as entirely separated from the real world (Agre, 1999). For example, Gibson (1984) defined cyberspace as a hallucination. In 2011, the cyberology blogger Nathan Jurgenson called this view, which was also reflected in policy and tech firm circles, "digital dualism" (Jurgenson, 2011).

Many researchers have questioned this paradigm of digital dualism and the clear-cut separation between the real and virtual world. Miller and Slater (2000) suggested framing the internet as an extension of the real world. They described digital spaces and technologies as "continuous with and embedded in other social spaces" (Miller & Slater, 2000, p. 5). Anthropologists such as Wilson and Peterson (2002, p. 456) and Hine (2015, p. 24) also strongly rejected the notion that there is a pre-existing distinction between the digital and the real world. This thesis supports this view and aims to highlight the offline consequences of online identity transformations.

As Tom Boellstorff argued in his paper “For whom the ontology turns”, even opposing the “digital” to the “real” world falsely equates the physical with the real and the digital with the unreal (Boellstorff, 2016, p. 387), which can “have devastating consequences for addressing the reality of the digital”. For example, losing money in online gambling or learning German online have real effects in offline settings. Likewise, some experiences can only become real by using digital means. Ginsburg (2012) demonstrated how digital media has allowed disabled people to transcend their physical limits and engage in activities that would otherwise be impossible through their virtual selves. In his “digital reality matrix”, Boellstorff (2016) therefore introduced four different categories: (1) the physical and real; (2) the digital and real; (3) the physical and unreal; and (4) the digital and unreal.

These different conceptions of the relationship between the virtual and the “real” have gone hand in hand with different methods applied to digital ethnography. Wilson and Peterson had noted already in 2002 that a coherent anthropological approach to studying digital cultures had yet to be developed (Wilson & Peterson, 2002). To date, two different approaches have dominated the field. The first one, developed by Miller (2011), uses traditional fieldwork to study online phenomena and their implications for individuals and collectives. For example, Miller conducted research on the use of Facebook through fieldwork on the Caribbean island of Trinidad, where existing local structures translated into Facebook networks.

The second approach is a fully online methodology that was introduced by Boellstorff (2008). For example, Boellstorff’s exploration of internet-based immersive virtual environments (Ves), such as the virtual shopping malls of Second Life, took place

purely online (Boellstorff, 2008). Using the avatar “Tom Bukowski”, Boellstorff conducted over two years of fieldwork in Second Life to study the relations between different players, including group bonds and conflicts, as well as sex, money, gender, and race dynamics. In an extensive 2012 handbook, Boellstorff and colleagues provided a practical overview of the different methods that can be used to conduct ethnographic research in online spaces (Boellstorff et al., 2012).

A third approach was introduced by Joshua Bluteau in 2019. Bluteau, who blends the real with the digital, describes his new methodology of “immersive cohabitation” as a move from participant observation to observing participant. His view is that the “digital landscape is no longer merely a research tool, but a field in its own right” (Bluteau, 2019, p. 5). For his study on fashion tailors and their clients in London, he developed an online presence on Instagram after realising the significance of the social media platform to his informants. For two years, he imitated what fashion influencers were doing, taking selfies every day to build an account that allowed him to gain access to this relatively closed digital field site. By combining traditional fieldwork with a self-immersion in the online community, he effectively managed to break the dichotomy of virtual and real-world fieldtrips. Likewise, Bonnie Nardi (2010) examined the multi-player game World of Warcraft by becoming a player herself and scraping thousands of chat logs, although she combined her online fieldwork with a one-month trip to Beijing to interview Chinese players of the game.

In the 2010s, many innovative studies were published, using a mixed approach of online and offline research. For example, McManus (2018) looked at how social media had transformed fandom among the Turkish football diaspora and argued that it has

given rise to a form of what he called “modern enchantment”. Based on his analysis of influential Facebook pages, as well as fieldwork and interviews with members of a German football club that supports the Turkish team Beşiktaş, McManus found that the possibility to transcend the limits of time and space in online networks has changed both the nature and the reach of fandom, enabling geographically distant diaspora communities to intensify their communication and cooperation (McManus, 2018, p. 763). MacKee (2016) explored the use of different dating apps by homosexuals in London, employing what he called a method of “deep hanging out”. His research drew on a mixture of online chats with people across the “media ecosystem of the gay subculture in London”, as well as real face-to-face dates and interviews (MacKee, 2016, p. 2).

In recent years, an increasing number of studies have focused on emerging online influencer industries and behavioural trends on digital platforms such as Tik Tok and Instagram. Digital ethnographic studies of the 2020s explored topics ranging from precarity and uncertainty in online video influencer communities (Glatt, 2021, 2022) to “bodyshaming” behaviours on Instagram (Fitria & Febrianti, 2020) and emerging metaverse gaming cultures (Han et al., 2021). The COVID-19 pandemic led to the field of digital ethnography expanding rapidly, as many researchers were forced to adapt and adjust their research designs (Johnson, 2022; Watson & Lupton, 2022). Simultaneously, there has been growing interest in studying online attitudes and behaviours related to the pandemic, such as the shift to remote education, as well as online discussions about health, vaccines, and COVID-related conspiracy myths (Arifka et al., 2022; Griera et al., 2022; Holdsworth, 2020). The Digital Ethnography Collective at the London School of

Economics (LSE), the OII, and the Digital Ethnography Research Centre hosted by Royal Melbourne Institute of Technology (RMIT) University have been among the institutions at the forefront of digital ethnographic work.

This doctoral project follows the tradition of Boellstorff in studying online environments and group bonds using an online-only methodology. However, it assessed the impact of online identities on real-world events by studying the online documents left behind by individuals who engaged in acts of real-world terrorism. Bluteau's innovative approach of "immersive cohabitation" was used in Chapters 4 and 5, which combines quantitative text analysis with self-immersion in a range of virtual groups. Moreover, the formulation of the thesis's hypothesis draws on several years of immersive cohabitation in online and offline settings of extremist communities for the books *The Rage* (Ebner, 2017) and *Going Dark* (Ebner, 2020).

Another source of inspiration for this project was the manifesto drafted by the Future Anthropologies Network at the conference of the European Association of Social Anthropologists (EASA) in 2014, which calls for a more future-oriented, networked, and unconventional approach to anthropological practice. Its authors argue that an increasingly complex present and uncertain future makes it necessary to reshape and complement the discipline of anthropology with experimental, speculative, transnational, and interventionist elements (Salazar et al., 2017). According to Arjun Appadurai, cultural anthropologists have a moral responsibility to engage in anticipatory and interventionist practices to emerging challenges (Appadurai, 2013). Similarly, James Ferguson suggested that anthropologists should play a role in contributing to "creating better human futures" (Ferguson, 2014). This is in stark contrast to Roy D'Andrade's

criticism of moral judgements in the discipline of anthropology to ensure that science and objectivity are not compromised (D'Andrade, 1995).

Ethical Considerations

While online extremism and radicalisation research is of significant public benefit and can help to inform law enforcement and security efforts to ultimately save lives, it also poses a unique challenge from an ethical and security perspective. Internet and data-driven research raises additional questions about privacy and data access. This thesis acknowledges the high moral responsibility of researchers in this particular field. Terrorism and extremism studies frequently require the use of innovative research methods to gain access to relevant channels in real-time and understand current and future trends in radicalisation. While moral judgements should not impede any of the thesis's research decisions or findings, reporting duties regarding dangerous and violence-inciting materials make moral judgement necessary.

This research project gave the highest priority to safeguarding vulnerable individuals, ensuring their anonymity and reducing the identifiability of any content that is analysed as part of this project. This research project followed guidance provided by the University of Oxford's Central University Research Ethics Committee (CUREC) Internet-Based Research (IBR) Best Practice Guidance (CUREC, 2019), the Association of Internet Researchers (AOIR, 2020), the University of Oxford's Safeguarding Code of Practice, and the University of Oxford's Code of Practice for the Safety of Social Researchers (University of Oxford, 2020a) to the greatest extent possible. Any personal details included in datasets were removed prior to the analysis and all usernames were

blurred in screenshots. All metadata or information that could help to identify individuals was removed. Data were only shared with the responsible authorities when there was a credible security threat that falls under the Prevent Duty. Security risks relevant to the Prevent Duty, as well as other illegal online activities, were flagged to the relevant security authorities and internet platforms. The use of deception was minimised and the provision of information about the project was carefully balanced with research safety considerations. The following list provides an overview of the most important ethics considerations for this project:

Vulnerable Participants: Adults at risk of radicalisation, as well as young people who are part of extremist forums, were at the heart of this research project. Due to the anonymous nature of the online forums, it was usually not possible to determine the age of users when scraping posts. As the goal of this project was ultimately to find better ways of safeguarding individuals from radicalisation, in particular adults "at-risk" and minors, it was important that these vulnerable groups were captured in the research. CUREC Approved Procedures 15 and 25 could be applied in this case, as the focus was not on educational settings and as the research covered the especially complex field of individuals at risk of radicalisation (University of Oxford, 2020b). From a safeguarding perspective, the project followed the University of Oxford's Safeguarding Code of Practice and the Code of Practice for the Safety of Social Researchers (University of Oxford, 2020a).

Anonymisation of Data: The focus of the project on extremism and radicalisation required touching on sensitive issues that are inherently part of the discussions happening in the forums that were analysed in the quantitative and qualitative analysis.

In line with the CUREC IBR Best Practice Guidance (CUREC, 2019) and the AOIR (2020), all data were anonymised and no personal details were published in any form. However, if users posted concrete threats, plans to plot terror attacks, or similar threats to national security, the content was flagged to the responsible authorities in line with the Prevent Duty (see below) and reported to the tech platform that hosted the account(s).

Prevent Duty: Any illegal activity, as well as medical or psychiatric conditions, discovered unintentionally during the course of the research was reported in accordance with the Prevent Duty and the Counter-Terrorism and Security Act (HM Government, 2015), the University of Oxford's Safeguarding Code of Practice, and the University of Oxford's Code of Practice for the Safety of Social Researchers (University of Oxford, 2020a).

Informed Consent: The collection of forum data was not based on informed consent as the data collection was only performed within publicly accessible forums, groups, and archived materials. This project recognises, however, that the perception of privacy from the perspective of research subjects may differ from the legal reality or terms of service. As such, it followed the recommendations of the CUREC IBR Best Practice Guidance (CUREC, 2019) and the AOIR (2020) to ensure the anonymity of all research subjects in all research outputs and publications, as well as guaranteeing that all case studies, quotations, or examples used in publications based on this thesis are non-identifiable.

Use of Deception: In line with the Approved Procedures for Deception of Adult Participants (University of Oxford, 2020b), as well as the recommendations of the British Psychological Society (2018), the following condition was applicable in this research

project: “Potential adverse effects to the researcher could arise from the research”. Due to security concerns, methods of deception were used in extremist online forums. However, deception was limited to the use of an avatar account that used a pseudonym and did not provide any identifiable information about the researcher’s identity.

Chapter 3: Towards a New Violence Risk Assessment Framework

This chapter presents a qualitative and quantitative comparative analysis of violent self-sacrificial manifestos, ideologically extreme (non-violent) manifestos, and ideologically moderate (non-violent) manifestos. Its aim is to distil patterns across texts that were published shortly before the respective authors committed acts of political violence. The thematic and linguistic characteristics found in terrorist manifestos were compared to the patterns identified in manifestos that were not followed by violence or extreme self-sacrifice by the authors.

Assessing the language and narratives in the written communications of terrorists, including manifestos associated with violent attacks, may help to establish a foundation for improving detection and prevention practices in the field of CT. Over the past decades, security and intelligence services, as well as social media companies, have tried to gain a better understanding of the communication patterns demonstrated amongst violent extremist and terrorists (U.S. Department of Homeland Security [DHS], 2018; Saltman, 2020; Zgryziewicz, 2018). Despite the high level of interest and stakes involved in establishing linguistic cues that can serve as early detection mechanisms to help prevent violence, there continues to be only a limited amount of published research in this area. This systematic manifesto analysis seeks to help address this gap in the research literature. Although the framework developed in this chapter does not have the predictive power to reliably single out every would-be terrorist, it could provide a useful tool within linguistic diagnostics, complemented with other forms of intelligence, to narrow down the potential targets of investigations. Given that security forces have

limited capacity to conduct investigations and there is often an excess of textual data available, producing effective linguistic analysis tools is crucial to helping CT units to act more efficiently. Most importantly, the tool that this thesis sought to develop is theory driven, based on extensive previous research on the psychological drivers of violent extremism.

The likelihood that someone puts their life on the line for a group has been linked to identity fusion, characterised by a porous boundary between personal and group identities (Swann et al., 2014). Numerous previous studies have shown that high levels of identity fusion predict willingness to fight and die for the group when it is imperilled (Whitehouse, 2018). Both non-state armed insurgents and conventional military groups in theatres of conflict in the Middle East have repeatedly demonstrated that fusion is a reliable predictor of actual violent self-sacrifice, as opposed to merely the endorsement of it (Tossell et al., 2022; Jong et al., 2015, Whitehouse et al., 2014, 2017). The same phenomenon has been observed with groups as diverse as Islamist fundamentalists in Indonesia (Kavanagh et al., 2019), farmers in Cameroon (Buhrmester et al., 2020), college fraternities in the US (Whitehouse et al., 2017), and football hooligans in Britain and Brazil (Newson, 2019; Newson et al., 2018). Despite the growing body of evidence linking fusion to violent self-sacrifice, its potential applications in counter-extremism efforts remain underexplored.

The comparative manifesto analysis is based on the fusion-plus-threat model (Whitehouse, 2018), which integrates previous empirical findings on group alignment and identity, psychological kinship, and parochial altruism to form a theory that makes specific predictions about the role of in-group identity fusion and out-group threat

perception as drivers of violent extremism. According to this theory, extreme ideological commitment or “radicalisation”, in the absence of fusion plus threat, would be much weaker predictors of violent extremism (Whitehouse, 2016). Predictive power is expected to be increased further when measures of relevant moderators of the relationship between fusion-plus-threat and violent extremism are available, such as violence-condoning norms (Whitehouse, 2018).

The model implies that reliable means of measuring identity fusion, perceptions of existential threat, and violence-condoning norms could help identify individuals at risk of carrying out terrorist attacks. Previous studies have used a wide variety of psychometric measures to assess these variables, but all rely on the willing participation of study participants. It is obviously impractical in most situations to persuade members of illegal groups or extremist networks, including those that condone or utilise violence, to provide honest answers to questions using measures of this kind. Consequently, the aim of this first part of the thesis was to explore whether linguistic markers found in actual terrorist manifestos can serve as reliable measures of fusion, threat, and willingness to condone the use of violence, as well as whether the predicted relationships are observed between these variables.

With the aim of identifying text-based predictors of terrorist actions, a qualitative and quantitative text assessment of 15 manifestos was conducted. The manifestos selected for analysis were produced by individuals who engaged in violent self-sacrifice, as well as ideologically extreme and moderate manifestos that were not followed by violent self-sacrifice by their authors. These included the manifestos of the terrorists behind the 2011 Norway attacks, the 2014 Isla Vista killings (US), the 2015

Charleston shooting (US), the 2019 Christchurch mosque attacks (New Zealand), the 2019 attacks in Halle (Germany), Poway (US), and El Paso (US), as well as publications by the National Socialist German Workers' Party (NSDAP) leader Adolf Hitler and the prominent jihadist revolutionary Sayyid Qutb. Furthermore, the selection featured manifestos of ideologically extreme but non-violent authors, such as the writings of Norwegian far-right blogger Fjordman, the Islamist ideologue Yusuf al-Qaradawi, and Marx and Engel's Communist Party Manifesto, as well as ideologically moderate manifestos by civil rights movement leader Martin Luther King Jr., feminist thinker Simone de Beauvoir, and climate activist Greta Thunberg.

Patterns in the narratives and language used in the manifestos were identified, distilled, and categorised to allow for a comparative analysis. A total of 100 sample phrases from the extracted manifesto content, which had been identified as representative of relevant coding categories, was subsequently collected. These phrases were then coded independently to investigate the coherence of the coding categories and the reliability of independent assessments (O'Connor & Joffe, 2020). More specifically, an intercoder reliability (ICR) analysis was conducted, first by two subject area experts and, subsequently, by 24 non-expert coders. The relevant narrative categories and related linguistic markers were subsequently quantitatively captured and statistically tested in an R-based NLP analysis of the manifestos.

Terrorism Prediction Models

Preventing acts of terrorism requires multi-dimensional risk assessment prediction models that take into account both environmental macro factors and individual micro-

factors. A growing number of theoretically informed, data-driven models to predict the risk of terrorism have been designed in recent years (Ding et al., 2017; Python et al., 2021). Many studies have focused on designing statistical models that seek to identify and quantify the effects of terrorism drivers in space and time (Nemeth et al., 2014; Python et al., 2018). Previous studies have used a combination of social factors (e.g. ethnic diversity, major drug regions, and populations density) and natural factors (e.g. night-time lights, population density, and topography) to simulate future conflict (Ball, 2012; Taylor et al., 2004). However, the underlying mechanisms are still not well understood (Ding et al., 2017).

The Institute for Economics and Peace created the Global Terrorism Index (GTI), which was one of the first large-scale attempts to measure the prevalence of terrorism and its impact across different geographies (Institute for Economics & Peace, 2019). In accordance with the Global Terrorism Database (GTD) and the National Consortium for the Study of Terrorism and Responses to Terrorism (START), the GTI defines terrorism as "the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation".

Drawing on existing literature on armed conflict and violent insurgency, Ding et al. (2017) used pioneering machine-learning methods to simulate the risk of terror attacks on a global scale. They trained their machine-learning models with data points from terrorist attacks that occurred between 1970 to 2015, and were able to predict the locations of terrorist attacks in 2015 with a success rate of 96.6%. While Ding et al.'s (2017) study may be helpful in informing the capacity planning of security services and

the geographic allocation of early prevention funds, the researchers aggregated data on a yearly level and therefore do not offer insights into short-term dynamics that would allow for early warning systems and timely law enforcement intervention. Python et al. (2021) followed this up with a new agent-based model using machine-learning techniques, which made it possible to predict the location of terrorist events on a fine spatial and temporal scale. Their model had higher predictive performance in short-term forecasting than Ding et al.'s (2017) approach.

This thesis does not look at macro factors and its goal is not to provide a holistic forecasting model that can predict the time and place of terrorist attacks. Instead, it seeks to make progress on the psychological risk assessment of individuals and groups by identifying and measuring recurring socio-psychological patterns in convicted terrorists and their expressions in written language. The aim of this research project is to inform prevention and intervention models about the human drivers and related linguistic markers that indicate a proneness to extreme forms of violence.

Several think tanks, tech firms, and government bodies have made progress in measuring violent extremist and hateful language deemed conducive to violence in virtual environments. For example, the ISD and CASM jointly created the "Hate Mapper", a tool that determines the concentration of anti-minority hatred on a borough level in different geographies, including the City of London and the State of Victoria in Australia. The Hate Mapper is based on lists of keywords that are seen as markers for anti-Semitic and anti-Muslim extremism (ISD, 2019b).

Though not developed for the purpose of examining terrorism only, it is worth mentioning Swann et al.'s (2014) "fight and die" scale as an alternative approach to measure willingness to engage in violent pro-group behaviour. Their original study, which introduced the measure, combined Swann et al.'s (2009) measure of the endorsement of extreme pro-group behaviours with two items that indicate willingness to die for their country (e.g. "I would sacrifice my life if it saved another country member's life") to form a more comprehensive "fight and die" index. The present project draws on previous literature to develop language-based means of identifying and measuring willingness to "fight and die", as well as the above-mentioned vulnerability factors in virtual groups.

As noted in the Introduction, neither academics nor policymakers have agreed on a universal definition of terrorism and violent extremism on a global scale. Against this background, it is not surprising that few measures to assess individual and group-based proneness to extreme violence exist to date. A reliable method of identifying would-be terrorists before they carry out attacks has not yet been found. Even if extreme ideology is frequently associated with terrorist violence, it is not a necessary condition, and its causal role has not been well established. Moreover, in recent years, a number of researchers have warned that the role of ideology in radicalisation has been exaggerated in counter-extremism approaches (Kundnani, 2015; Powell, 2016). As argued in Chapter 2, the effort to link terrorist and extremist profiles has been inconclusive (Gill et al., 2014; Neumann, 2017). As a result, researchers studying radicalisation have increasingly turned their attention to potential psycho-social drivers that bridge personal grievances and group narratives.

Prevention and deradicalisation approaches have evolved in parallel with emerging evidence-based research. Policymakers and frontline practitioners are now increasingly aware of psychological drivers that have emerged as consistent patterns in radicalisation pathways. In the post-9/11 era, the idea that ideology is the starting point from which individuals mechanically move towards violent extremism has become known in policy circles as the "conveyor belt" theory. This doctrine gained prominence in the UK when former Prime Minister David Cameron used the term in a 2011 speech at the House of Commons to call for stronger action to tackle non-violent Islamist extremism (Cameron, 2011). In recent years, radicalisation prevention policies have undergone notable changes, mirroring both the theoretical shifts towards multi-dimensional explanation models as well as integrating new evidence emerging from psychological research on radicalisation (for a more detailed account, see Koehler and Horgan, 2016).

A number of transnational organisations, including the European Parliament (2014), the UN (United Nations Development Programme [UNDP], 2016), the OSCE (Neumann, 2017), and the Human Rights Council (Emmerson, 2016), have made efforts to do justice to the complexity and diversity of pathways to radicalisation. For example, an OSCE report by Peter Neumann of the ICSR at King's College named five recurring factors: grievances; needs; ideas; people; and violence (Neumann, 2017, pp. 17–18). Magnus Ranstorp, who helped to establish the European Union's flagship radicalisation research programme Radicalisation Awareness Network (RAN), described radicalisation as a "multifaceted combination of push-pull factors involving a combination of

psychological factors, political grievance, religious motivation and discourse, identity politics and triggering mechanisms" (Ranstorp, 2010, p. 6).

The US national terrorism risk assessment guide emphasises motivational factors as indicators of the threat to violent mobilisation. While it cites ideology, relationships, and intent as important motivation indicators, it fails to mention the role of identity and group cohesion in the radicalisation process (National Counterterrorism Center [NCTC], FBI, & DHS, 2021). The New York State Police Department treats radicalisation leading to jihadist terrorism as a process composed of four distinct phases that echo Sageman's model: pre-radicalisation; self-identification; indoctrination; and jihadisation (Silber & Bhatt, 2007). Sageman's credentials as a former CIA case officer and the first scholar in residence at the New York Police Department has no doubt contributed to the influence of this model in US policy and security circles.

The UK government highlights the role of identity crises in bridging psychological and ideological contributions to radicalisation. The Home Office's framework, which informs all top-down prevention and intervention programmes in the UK, depicts radicalisation as a cocktail of personal grievances, identity crises, exposure to charismatic recruiters, and their extremist narratives and ideologies (HM Government, 2010). While the British Prevent programme initially had a strong focus on Islamist ideologies, attracting much criticism for its disproportional focus on religious and cultural minority communities, it has increasingly emphasised the role of overarching patterns in social processes and identity transformations in recent years.

Building on existing research on radicalisation pathways, today's Channel programme (the UK's radicalisation referral programme) integrates a multitude of psychological factors identified as relevant in the radicalisation process (HM Government, 2015). The Channel programme is an essential part of the British CT strategy CONTEST and at the heart of its "Prevent" pillar. Since the Prevent Duty was introduced in the UK in 2015, teachers, prison guards, and other frontline workers are statutorily obliged to refer individuals they believe to be at risk of radicalisation to the Prevent programme (HM Government, 2015). Following an initial confidential screening of referred radicalisation cases by the police, referrals may then be forwarded to the multi-agency "Channel panel", which consists of local-authority representatives, education and health practitioners, and radicalisation experts. The Channel committee's decision on whether referred radicalisation cases should receive support via the Channel's one-on-one deradicalisation programme is informed by the Vulnerability Assessment Framework, which covers a mix of psychological and environmental risk factors on an individual level: (1) engagement with a group, cause, or ideology (e.g. feelings of grievance and injustice, perceived threat, need for identity or belonging, desire for status, desire for political or moral change, etc.); (2) intent (e.g. "us versus them" thinking, dehumanisation of the enemy, violence justification, etc.); and (3) the capability to cause harm (e.g. individual knowledge, skills, and competencies, access to networks, funding, or equipment, etc.) (HM Government, 2012).

Despite recent developments in policy approaches to radicalisation, there is potential to further enhance prevention and intervention models based on evidence from psychological research. While there are many ways to achieve this, the following

section focuses on the future of early warning systems that can assist frontline workers and security services in identifying recurring risk factors of violent extremism.

Using Language to Detect Would-Be Terrorists

Government policies, legal frameworks, and intelligence investigations tend to be based on explicit online threats or expressions of support for terrorism in an effort to reduce the danger of violent attacks and crimes (Council of Europe, 2022; FBI, 2022; HM Government, 2015). Verbal threats, however, are not a reliable predictor of actual violence as many users make empty threats (Fein & Vossekuil, 2000; Ovide, 2021). Consequently, the challenge is to establish a more robust set of predictors of violence based on the analysis of more fine-grained patterns of language use. The aim of this chapter is therefore to design and test a theoretically grounded approach to assessing the risk of violence based on language that indicates strong motivations to engage in extreme forms of pro-group action, regardless of whether acts of violence are specifically threatened or not. The goal is not to offer a predictive model at an individual level but rather working to build the foundations of a model and associated linguistic toolkit that identifies recurrent factors associated with subsequent violence.

Several studies have performed language-based content analysis of terrorist materials (Pennebaker & Chung, 2008; Smedt et al., 2018; Qui, 2020, Smith et al., 2008), although they were not specifically focused on the manifestos published online by future terrorists. For example, Qui (2020) published an empirical analysis of 79 official ISIS videos, examining the prevalence and evolution of speech acts in the timeframe 2014–2017. The conclusion was that threatening speech featured most prominently in

the analysed videos. Directive, expressive, and assertive speech was also commonly used, reflecting the terrorist group's aim to mobilise its audiences and achieve a diplomatic effect in addressing real-world events. Likewise, Smith et al. (2008) completed a quantitative content analysis, comparing the value references, motive imagery, and integrative complexity expressed in two Al-Qaeda-affiliated terrorist groups and two non-terrorist control groups. The study's statistical analysis showed that the two terrorist groups used more expressions of positive morality, religion, and aggression value references to describe the in-group, while linking their enemies to negative religion values (e.g. labelling them as infidels).

Systematic linguistic analyses of violent and non-violent manifestos are scarce. There have been several publications that have analysed the themes, narratives, and language of individual terrorist manifestos, but these previous works did not perform a comparative analysis (Barnett, 2015; Beutel, 2019; Chebrolu, 2020; Gardell, 2014). The most extensive linguistic analysis of terrorist manifestos was conducted by researchers at Uppsala University in 2016 (Kaati et al., 2016). Their study focused on "leakage warning behaviour" (or simply "leakage"), which describes the intentional or unintentional signalling of planned violence in public or non-public communications. The purpose of the leakage that was observed in the study with the perpetrators ranges from attention-seeking, coping with anxiety related to the impending act, or the desire to memorialise one's deed beyond one's death.

The Uppsala group used the text-analysis tool Linguistic Inquiry and Word Count (LIWC) to assess the patterns in written documents authored by 10 different lone terror offenders prior to their engagement in violence. Words were coded in psychologically

meaningful categories, yielding eight indicators of the drives and emotions that preceded violent attacks: *the use of big words*, i.e. words longer than six letters have been found to be indicative of psychological detachment (e.g. degeneration, resistance, overconsumption), *the use of personal pronouns*, in particular third-person plurals (e.g. they, their, them), expressions of emotions, including the ratio of *positive emotions* (e.g. happy, good, pretty) to *negative emotions* (e.g. hate, enemy, worthless) and *anger words* (e.g. hate, kill, annoyed), words related to *friends* (e.g. buddy, neighbour, brother), *certainty* (always, never), and *power* (e.g. superior, bully, weak). The identified communication features were also compared with those found in natural conversations (Kaati et al., 2016).

The results of this computerised textual analysis of manifestos offer an illustration of its use in exploring the language and narrative patterns in terrorist manifestos. One significant limitation of the Uppsala method, however, is that the approach was not well-grounded theoretically and did not test predictions based on previous evidence-based research. For instance, although the authors highlight the category "friends" as an important one that emerges from their analysis, and even quote "brother" and "buddy" as example terms, they do not reference any previous work on psychological kinship or fusion theory. Likewise, the study cites linguistic markers of cognitive processes, in particular "the simplification of complex matters", without linking its findings to established measures of cognitive complexity, such as integrative complexity (Hiel & Mervielde, 2003).

Another limitation of Kaati et al.'s (2016) study stems from the timing of the publication. As the analysis was completed in 2015, it does not include the manifestos

of more recent high-profile lone actors, such as those that were posted in the run-up to the attacks in Christchurch, Poway, Halle, and El Paso. The authors furthermore offer word categories in their publication but do not provide lists of specific terms that can be used for NLP analysis. The authors similarly acknowledge that future work looking at linguistic cues should seek to include a range of manifestos for analysis, as well as comparing differences individually with their categories as a potential foundation.

Another important and more recent contribution to the analysis of the content of terrorist manifestos was made by Jacob Ware of the International Centre for Counter-Terrorism (ICCT), who carried out a qualitative analysis of the manifestos of extreme-right terrorists (Ware, 2020). According to Ware, recurring themes in the manifestos were an alleged "clash of civilisations" between races or religions, the importance of the preservation of European culture, the political climate, and the portrayal of terrorism as both self-defence and a last resort. While this analysis focused on thematic patterns, it did not assess technical linguistic characteristics. The same can be said of Florian Hartleb's compelling systematic analysis of single-actor terrorists, which zooms in on the individual biographies of prominent terrorists. His book *Lone Wolves* provides a deep analysis of the motivational patterns found in individual pathways of radicalisation towards violence and draws conclusions about overarching trends (Hartleb, 2020). The fusion-based violence risk assessment framework presented in this thesis incorporates the learnings from these previous studies but focuses on the text-based detection of linguistic markers associated with relevant narrative patterns.

In contrast to the previous comparative manifesto-related studies, which were not clearly theoretically grounded, the following analysis presented in this thesis uses

the recently developed fusion-plus-threat model (Whitehouse, 2018) as a conceptual basis for the qualitative and linguistic analytical approach. In so doing, it seeks to provide a potentially more generalisable approach to identifying linguistic markers in narratives that are capable of predicting violent as opposed to non-violent outcomes. This chapter's aim is to build on the valuable existing research outlined above and trigger wider discussion concerning the content and linguistic cues in violent versus non-violent manifestos, rooting the analysis in an evidence-based framework.

As mentioned in Chapter 1, fusion is thought to be more effective at motivating extreme pro-group behaviour than group identification (Whitehouse, 2018). A long tradition of research in the field of social identity examining how people bond with groups indicates that group identification is a depersonalising process (Rosenberg, 1987). Thus, when a person strongly identifies with a given group identity and that identity is made salient, their personal identity becomes less salient, implying a hydraulic relationship between the personal and group identities (Hogg & Reid, 2006; Swann & Buhrmester, 2015). However, more recent research has shown that identification is primarily based on sharing identity markers with other members of a group, such as beliefs and practices, while identity fusion derives from sharing traits or experiences that define the personal self, such as life-defining experiences as in the case of persecuted or embattled groups (Gómez et al., 2011; Swann et al., 2009).

Consequently, a strong conflict between self-interest and group needs can generate a dilemma that may more readily be resolved in a self-preserving manner when identification rather than fusion is the dominant form of group alignment and depersonalisation is not evident. Fused individuals, by contrast, due to the synergistic

nature of the self and group bonds, appear more willing to take extreme actions to protect the group as if it is equivalent to their personal self (Swann et al., 2010). While strong group identification often occurs in tandem with high levels of identity fusion, the theoretical framework presented here based on previous research (Swann et al., 2014; Whitehouse, 2018; Whitehouse et al., 2017) leads to the prediction that, in the presence of threat, fusion rather than identification is more often a potent driver of violent extremism.

It is theoretically and methodologically important to be able to distinguish between the two forms of group alignment when attempting to explore signals of willingness to conduct personally life-threatening acts of violent out-group hostility, rather than merely endorsing such acts philosophically or issuing hollow threats. An important caveat is that, according to the proposed framework, fusion alone is not sufficient to motivate out-group violence. This thesis also does not claim that the fusion-plus-threat model is the only way of explaining violent self-sacrifice, since it is possible that other motivators may be sufficient, ranging from violence-endorsing group norms to psychosis and suicidalism (Lankford, 2011, 2014; Sheehan, 2014). Nevertheless, to the extent that the fusion-plus-threat model explains much of the variation in this domain, it is vital to develop a set of diagnostic tools for early prevention.

A plausible diagnostic marker for high levels of fusion in written or verbal form is the presence of language emphasising relations of kinship or familial ties (e.g. siblingship) among group members. From the outset, research on identity fusion has emphasised the link with family ties and feelings of brotherhood and sisterhood in fused groups (Swann et al., 2014). Studies comparing pathways to fusion among large samples

of identical and fraternal twins indicate that perceptions of shared biology, as well as shared experience, contribute to identity fusion (Vázquez et al., 2019; Whitehouse et al., 2017). Studies among survivors of atrocities also show that feelings of psychological kinship mediate the relationship between fusion and self-sacrificial behaviours (Buhrmester et al., 2014). The pervasive use of metaphors of kinship when talking about the in-group and its members can thus indicate heightened levels of identity fusion, and the associated willingness to engage in extreme pro-group behaviour, including violence and terrorism (Whitehouse & Lanman, 2014). There is therefore a compelling theoretical rationale for treating expressions of kinship relatedness among members of groups as indicators of high levels of fusion.

It is important to acknowledge the rhetorical aspects of manifesto writing. The use of specific linguistic proxies could be construed as a conscious rhetorical choice that serves to performatively construct and/or reinforce a collective group identity. In line with the theory of language outlined in the Introduction, this thesis argues that the use of kinship language, whether rhetorically chosen or unconsciously expressed, is an indicator of the author's level of identity fusion. The language used by the manifesto writers may be seen as a messenger of psychological drives or as an active effort to strengthen the respective in-group the author is fused with. From a teleological perspective, the question whether a manifesto author's words were strategically chosen or the outcome of underlying psychological processes is not expected to impact the usefulness of the kinship markers to draw conclusions about the author's level of fusion.

Qualitative Comparative Manifesto Analysis

This manifesto analysis applies the predictions of fusion theory by exploring whether the language found in different types of manifestos could be used as a diagnostic tool in the prediction of violent extremism. In the process, it is important to disambiguate linguistic markers of fusion and identification, as well as to clarify which clusterings of other relevant variables demonstrate mediating connections between group bonding indicators and violent outcomes, and which do not. Using the fusion-plus-threat framework for predicting violent self-sacrifice (Whitehouse, 2018), the hypothesis is that a higher degree of identity fusion language, coupled with potential moderator variables (detailed in Figure 2), increases the likelihood that individuals will carry out violent pro-group action. This means that manifestos that have led to acts of self-sacrificial violence would be predicted to include more fusion language than those in the categories "ideologically extreme non-violent" and "moderate non-violent".

Importantly, the fusion-plus-threat framework, as the name indicates, argues that fusion alone is unlikely to predict violent self-sacrifice without threat or relevant group values. As outlined in the Introduction, this thesis aims to capture other potential variables of interest, including the proposed moderators of the pathway from fusion to violence. Therefore, the data were coded for linguistic indicators of: (1) group alignment; (2) out-group entitativity; (3) out-group threat; (4) violence-condoning norms; and (5) violence threats. See the later "Narrative Coding" section (Table 3) for full narrative descriptions and Figures 1 and 2 for visual representations of the proposed relationships between relevant variables.

Calls to violence were treated as distinct from violence-condoning norms, since adhering to norms condoning violence does not necessarily mean that one considers violence the best method of securing the group's present interests. Moreover, it is important to recognise that merely threatening violence is no guarantee that one will actually engage in violence. While many cybercrime units and intelligence agencies traditionally focus on detecting explicit threats of violence and support expressed for terrorism in their prevention and intervention efforts (Council of Europe, 2022; FBI, 2022; HM Government, 2015), this thesis puts forward the argument that this variable alone may not represent a reliable predictor of actual violence. In online communities, many calls to violence remain no more than empty threats or efforts at sarcasm, while an increasing number of today's violent extremists have become more cautious in their use of threatening language to circumvent detection mechanisms and removal policies (Ebner, 2020; Keen et al., 2020). The following fusion-plus-threat-based framework thus aims to serve as an alternative or complementary prediction model that can potentially identify higher risk combinations of language. Tracking the direct calls to violence in terrorist manifestos, however, remains important as it enables examining how this variable interacts with other factors and clarifying when this variable correlates with actual violence.

Manifesto Selection

A total of 15 manifestos were selected for the analysis, on a scale from violent self-sacrificial to non-violent, to enable comparative analysis across different types of writings. The selected ideologically extreme (violent and non-violent) manifestos differ in ideological leanings and cover anti-Muslim, anti-Semitic, misogynist, and jihadist

documents, while the moderate (non-violent) manifestos cover prominent feminist, environmental, and anti-racist texts. All assessed text-based documents are comparable in three aspects: their declaration of political or ideological views; the expressed urgency of their political or ideological message; and their targeting of public audiences. While all of the authors called for radical change, the tactics they pursued to reach them varied widely following the publication of their statements. The total number of analysed manifestos was limited to 15 due to the limited availability of terrorist manifestos, access restrictions to such documents, as well as researcher time considerations. A total of over 4,000 pages were manually reviewed.

A range of terrorist manifestos were included to capture relevant high-profile violent attacks that occurred between 2011 and 2021, where the perpetrators had left behind manifestos or extensive letters about their motivations. The authors of the selected terrorist manifestos can be described as lone actor terrorists who committed acts of political violence on their own and were not officially part of traditional terrorist groups. As Jacob Ware observed in a recent report, the use of manifestos by terrorists has increased in recent years (Ware, 2020). However, it is worth noting that manifestos are more commonly used by extreme-right terrorists than by violence perpetrators of other ideological leanings. While ISIS frequently claimed responsibility for attacks ex-post via their Amaq News Agency or its propagandists, terrorist manifestos similar to the ones examined in this chapter are predominately a far-right phenomenon.[2]

[2] This observation was made based on a systematic search of academic articles, news reports, primary source archives, and a consultation with the founder of the jihadist primary source website Jihadology, Aaron Y. Zelin.

One explanation for the discrepancy in the use of manifestos by perpetrators of different ideological backgrounds may lie in the frequency of copycat terrorism, where a perpetrator helps to inspire the next through their manifestos and attacks, which is more prevalent among the extreme right. When the Norwegian terrorist Anders Behring Breivik published his 1,500-page manifesto before launching his attack on 22 July 2011 he initiated a phenomenon of posting documentation prior to far-right attacks that cannot be observed with jihadist perpetrators (Ware, 2021). A *New York Times* investigation found that the majority of white terrorists since 2011 referred to previous right-wing attackers and their manifestos in their written communications (Cai & Landon, 2019). Indeed, during the drafting of this manuscript, there was another mass shooting in Buffalo, New York, that resulted in 10 fatalities. The accused shooter, Payton S. Gendron, composed a 180-page manifesto which was reported to have plagiarised about 57% of the ideological sections from previous sources, especially the Christchurch shooter manifesto (Ihler, 2022).

While jihadist attackers have often been found to be inspired by political and theological manifestos written by Islamist ideologues, they do not tend to leave behind manifestos themselves (Ayad et al., 2021). However, violent self-sacrifice is not limited to terrorists. Although not classifiable as terrorist manifestos, Hitler's *Mein Kampf* and Sayyid Qutb's *Milestones* are writings by authors who subsequently participated in the planning of acts of extreme violence. These writings were included to also assess the narratives and language used by perpetrators of high-profile forms of political violence that do not fall under the category of lone actor terrorism.

The selected terrorist manifestos were all written in the 2010s and directly followed by violence. However, the other manifestos were taken from different timeframes and cultural contexts. Given the thesis postulates a set of psychological drivers of violent extremism that operate independently of cultural and historical contexts, the time and space in which the selected texts were written should not change the outcome of the analysis. As mentioned above, Hitler and Qutb were included as historical references of extreme violence associated with different ideologies. As no jihadist terrorist manifestos could be found due to the reasons stated above, Qutb's text represents an important case of Islamist extremist associated violence. The aim for the non-violent control datasets was to capture high-profile manifestos of individuals in different cultural contexts who never resorted to violence during their lifetime. The analysis therefore includes a range of historically influential ideologically extreme and ideologically moderate pieces of political writing. In the cases of the feminist and civil rights movement it can be safely stated that neither Simone de Beauvoir nor Martin Luther King Jr used means of extreme violence during their lifetimes. The same is true for the authors of the selected ideologically extreme non-violent manifestos.

Due to its relatively young age, the environmental movement does not have a similar historic figure whose writings can be analysed in this context. Even though Greta Thunberg inspired a peaceful, non-violent movement and has herself never been engaged in any acts of violence since the inception of the Fridays4Future movement in 2018, the inclusion of her climate action manifesto in the non-violent control group is only based on the status quo in 2023. Environmental manifestos such as the *Earth Charter* (2000) or the Extinction Rebellion's *Declaration of Rebellion* (2018) were not

suitable for the analysis, as they were published by organisations rather than individual authors. However, alternative texts could have included *Deep Ecology for the Twenty-First Century* by George Sessions (1995) or *Ecofeminism* by Maria Mies and Vandana Shiva (1993).

The following list provides an overview of the analysed manifestos:

Anders Behring Breivik, Oslo and Utøya (2011): Anders Behring Breivik committed one of Europe's deadliest terrorist attacks. On 22 July 2011, he first carried out a car bomb attack against a government building in Oslo, killing eight people and injuring at least 110 others. A few hours later, he went to the Workers' Youth League summer camp on Utøya island, where he shot dead 69 people while dressed as a Norwegian police officer. On the day of his attack, Breivik published a 1,500-page manifesto "2083 – A European Declaration of Independence", which outlined his ideological beliefs and reasons for the attack (Breivik, 2011). As pointed out by Kundnani (2012b), large parts of the manifesto can be considered a compilation of texts copied from US far-right websites. It used extracts from FrontpageMag, JihadWatch, Gates of Vienna, and the blogs of famous counter-jihadists such as Bat Ye'or, Fjordman, Robert Spencer, and Serge Trifkovic. Nevertheless, the manifesto provides insights into the thinking and motivations of Breivik. In the opening pages, the convicted Norwegian terrorist claims that it took him several years to complete the manifesto. He discusses the so-called Eurabia conspiracy theory and his perceived threat of the Islamisation of Europe and outlines strategies for armed resistance and communication tactics among "cultural conservatives" against the "cultural Marxists". He names political correctness, cultural Marxism, radical feminism, critical theory, and globalised capitalism as the culprits that "have destroyed every

defensive structure of European society which has laid the foundation for the Islamisation of Europe.” He speaks of an inevitable “European civil war”, explains how to deal with traitors of different categories (“hardcore Marxists, cultural Marxists, suicidal humanists/career cynicists and capitalist globalists”) and provides instructions for preparing explosives, using criminal networks, and launching attacks.

Elliot Rodger, Isla Vista (2014): On 23 May 2014, the 22-year-old US citizen Elliot Rodger committed a series of attacks against students in Isla Vista near the campus of the University of California. He killed six people and injured 14 others in a combined gunshot, stabbing, and vehicle attack before committing suicide. The manifesto Rodger left behind has the character of an autobiography that describes his radicalisation towards misogyny (Rodger, 2014). It outlines what he considers to be his most important life experiences that shaped him as a person and motivated him to carry out the terror attack. In his manifesto, Rodger talks of travelling with his parents at a young age, of how his parents got divorced, and of how he was repeatedly bullied for being the weird, shy kid in school. He describes his social rejection, loneliness, and jealousy towards boys that were more popular with girls, as well as his obsession with World of Warcraft. Compared to other terrorist manifestos, Rodger spoke much more in “I” versus “Them” rather than an “Us” versus “Them” divides. He views the battle as one where he is opposed to the rest of the humanity. The manifesto describes a few personally transformative experiences that Rodger blames for his hatred towards girls and women.

Dylann Roof, Charleston (2015): The white supremacist US citizen Dylann Roof chose the Emanual African Methodist Episcopal Church in Charleston, South Carolina, for his mass shooting on 17 June 2015. Nine African Americans were killed, making the Charleston

church shooting one of the deadliest attacks on a place of worship in US history. Dylann Roof published a manifesto that detailed his hatred of ethnic minorities, including Blacks, Jews, Hispanics, and East Asians. He describes “modern American patriotism” as “an absolute joke” and justifies his own use of violence, writing that he is left with “no choice” (Roof, 2015).

Brenton Tarrant, Christchurch (2019)*:* On 15 March 2019, Brenton Tarrant carried out two consecutive mosque attacks in Christchurch, New Zealand, killing over 50 Muslim worshippers. Brenton Tarrant livestreamed his attack from a first-person ego-shooter angle with a GoPro camera attached to his head and released an extensive 74-page manifesto online on the far-right image board 8chan. In his manifesto, Tarrant makes excessive use of insider vocabulary from far-right extremist online sub-cultures. Thematically, the document describes turning points, or transformative experiences, that led to Tarrant’s radicalisation (Tarrant, 2019). The author names jihadist attacks, as well as political and demographic developments, as notable factors on his journey towards violence. Of particular importance was the so-called “great replacement” conspiracy theory, which describes the idea that whites are being gradually replaced by non-whites and face the danger of extinction (Macklin, 2019).

John Earnest, Poway (2019): On 27 April 2019, the last day of the Jewish Passover holiday, John Earnest attacked a synagogue in Poway, California. He killed one person and injured several others, including the synagogue’s rabbi. Although Earnest cited the Christchurch shooter as inspiration and quoted the idea that whites are being gradually replaced, his hatred was not primarily directed at Muslims. Instead, he vowed to kill Jews, whom he held responsible for the alleged genocide against “the European race”.

Like Tarrant, Earnest published his open letter on 8chan. His text documents his own family history, as well as the reasons for his hatred of non-whites and Jews. It includes a range of insider terms and jokes from online far-right subcultures (Earnest, 2019).

Stephan Baillet, Halle (2019): The Halle attack happened on 9 October 2019, when the heavily armed 27-year-old Stephan Baillet attempted to enter a synagogue in Halle, Germany. After Baillet failed to make his way into the synagogue on the Jewish holiday of Yom Kippur, he opened fire on by-passers and at a nearby kebab shop. He fatally shot two people and injured two others. Prior to his attack, Baillet uploaded a manifesto to the online message board Meguca, which is loosely related to 4chan. The focus of the document is on his weapons, equipment, and attack plan. However, he also outlines his objectives and lists "achievements" for different targets and killing tactics in reference to video game culture (Baillet, 2019).

Patrick Crusius, El Paso (2019): On 3 August 2019, 23 people were killed and 23 injured when the 21-year-old US citizen Patrick Crusius carried out an attack at a supermarket in El Paso, Texas. The shooting was the deadliest attack on Latinos in modern US history. Imitating both Tarrant and Earnest, Crusius cited the "great replacement" theory as inspiration and uploaded his manifesto to 8chan. His writings outline his anti-immigration views and concern over the birth rates of white Americans. He writes both about his political and economic reasons as well as personal grievances (Crusius, 2019).

Adolf Hitler, *Mein Kampf* (1925): Hitler wrote *Mein Kampf* during his imprisonment, following his failed *coup d'état* in 1923. After Hitler's ascent to power in 1933, the book became a nation-wide bestseller in Germany. The 720-page document is both an

autobiography and a political manifesto. In the beginning, *Mein Kampf* chronicles Hitler's childhood and youth, from early ambitions of becoming a painter, the struggles with his own father who wanted him to become a government official, and his parents' deaths to his first encounters with Jews (Hitler, 1981). It exposes his personal grievances and describes his path towards radical anti-Semitism. Beyond this, *Mein Kampf* also introduces concepts such as "*Lebensraum*" ("living space") and proposes strategies that would later inspire the NSDAP's laws, propaganda, and violence. Historians have pointed out that the genocidal nature of Hitler's thinking was already evident in *Mein Kampf* (Kershaw, 1999).

Sayyid Qutb, *Milestones* (1964): The book *Milestones* by the 20th century Islamist ideologue Sayyid Qutb is a manifesto for forming a true Muslim society. Qutb was an Egyptian philosopher (who is often called "the father of Salafi-jihadism") and a leading member of the Muslim Brotherhood in the 1950s and 1960s. In 1966, the author was sentenced to death by hanging for plotting the assassination of Egyptian President Gamal Abdel Nasser. In his lifetime, he published over 20 books, with *Milestones* being one of his most influential. In *Milestones*, he writes about the necessity for violence in order to free humanity from "*jahiliyya*" (ignorance). He asserts that the Islamic vanguard must spread the religion "boldly, clearly, forcefully, without hesitation or doubt" (Qutb, 2006, p. 112). His writings served as an intellectual inspiration for several jihadist terrorist groups, including Al-Qaeda and ISIS. For example, Qutb was quoted in the 9/11 Commission Report as having played a significant role in the radicalisation of the perpetrators behind the attacks on the Twin Towers (National Commission on Terrorist Attacks upon the United Sates, 2004).

Karl Marx and Friedrich Engels, *Manifesto of the Communist Party* (1848): The Communist Manifesto is a pamphlet co-authored by Karl Marx and Friedrich Engels. The document, which was commissioned by the Communist League and published in the UK at the beginning of the revolutions of 1848 deals with the history of class struggles and explains the need for communism in order to eradicate systemic injustices and inequalities. It details the demands and principles of communism and offers a discussion of different forms of socialism and their relationship with communism (Marx & Engels, 2004). The manifesto was one of the most influential political works, inspiring both violent revolutions and violent dictatorships. The manifesto contains incitement to violent action such as "The Communists openly declare that their ends can only be attained by the forcible overthrow of all existing social conditions". However, some historians argue that Lenin and Stalin explicitly rejected the Marxist view that peaceful revolution is possible and reinterpreted the Communist Manifesto in favour of a more violent approach than envisioned by Marx (Hook, 1973, p. 275).

Yusuf al-Qaradawi, *The Lawful and Prohibited in Islam* (1960): *The Lawful and Prohibited in Islam* is one of over 120 books published by the Egyptian scholar Yusuf al-Qaradawi. As the name suggests, *The Lawful and Prohibited in Islam* is a guide for Islamic life that denounces Western behaviour and lifestyle choices (al-Qaradawi, 1960). Al-Qaradawi is based in Qatar and has been labelled a "moderate Islamist ideologue" (Baker, 2003). Al-Qaradawi is an intellectual leader of the Muslim Brotherhood and known for his condoning of Palestinian suicide bombings against Israelis. In 2008, he was refused entry into the UK and, in 2012, he was banned from travelling to France. He was the founder

of the global online platform IslamOnline and chairman of the International Union of Muslim Scholars (Halverson, 2017).

Fjordman, *Defeating Eurabia* (2008): *Defeating Eurabia* is a self-published book by Peder Are Nøstvold Jensen, a Norwegian far-right blogger who writes under the pseudonym Fjordman (Fjordman, 2008). Anders Breivik's manifesto quoted Fjordman's articles and was inspired by the idea of Eurabia, the conspiracy theory that suggests Europe is facing a gradual take-over by Muslim immigrants. *Defeating Eurabia* is a compilation of Fjordman's blog entries on far-right websites such as Gates of Vienna, Jihad Watch, Atlas Shrugs, and The Brussels Journal. It focuses on criticising mass immigration and the alleged Islamisation of Europe.

Simone de Beauvoir, *The Second Sex* (1949): *The Second Sex* ("*Le Deuxième Sex*", originally published in French) by the French existentialist philosopher Simone de Beauvoir is one of the most influential feminist works of the 20th century. In the book, de Beauvoir asks the question "what is woman?" She criticises how females are societally regarded as "the other" relative to males and are raised to be women, which leads to their submission and reproductive slavery. The book offers a discussion of biological and societal differences between the sexes, as well as a critique of the treatment of women throughout history (de Beauvoir, 2011). *The Second Sex* played an important role in inspiring second-wave feminism in the 1960s and peaceful women's rights protests such as the French Women's movement (Monteil, 1997, p. 7).

Martin Luther King Jr., *I Have a Dream* (1963): King's famous "I Have a Dream" speech turned the social activist into the most influential intellectual leader of the civil rights

movement from the mid-1950s. The speech played a key role in giving rise to new forms of peaceful protest and civil disobedience. It was a plea for more justice and equality for African Americans, but King also shared his own dream of what a more just society would look like (King Jr., 1963).

Greta Thunberg, *Our House is Still on Fire* (2020): In 2018, the Swedish teenager Greta Thunberg founded the youth-led street movement Fridays4Future, which by 2021 had reached 14 million followers across the world. Greta Thunberg gave her speech "Our House is Still on Fire" in front of political leaders at the World Economic Forum in Davos in 2020. The speech highlights the extent and urgency of the climate crisis and appeals to leaders in business and politics to take immediate action (Thunberg, 2020). Thunberg's speeches and writings have inspired school strikes, peaceful protest, and days of action in thousands of cities on all continents (Fridays4Future, 2022).

Manifesto Coding

All the manifestos used in this study were coded on a spectrum from violent to non-violent texts, which reflects the different pathways based on the fusion-plus-threat model. More specifically, the analysis distinguishes between three types of manifestos (see also Table 2):

(1) Violent self-sacrificial manifestos: texts written by authors who committed acts of self-sacrificial violence.

(2) Ideologically extreme (non-violent) manifestos: political texts written by authors who did not commit acts of self-sacrificial violence but adhered to extremist group ideologies.

(3) Moderate (non-violent) manifestos: political texts written by authors who did not commit acts of self-sacrificial violence and did not adhere to extremist group ideologies.

Table 2: Manifesto Coding

Author	**Description**	**Length**	**Coding Category**
Anders Behring Breivik	Manifesto of the Norway attacks, 2011	1,500 pages	Violent self-sacrificial
Elliot Rodger	Manifesto of the Isla Vista killings, 2014	140 pages	Violent self-sacrificial
Dylann Roof	Manifesto of the Charleston shooting, 2015	5 pages	Violent self-sacrificial
Brenton Tarrant	Manifesto of the Christchurch mosque attacks, 2019	74 pages	Violent self-sacrificial
Stephan Baillet	Manifesto of the Halle synagogue shooting, 2019	11 pages	Violent self-sacrificial
John Earnest	Manifesto of the Poway synagogue attack, 2019	11 pages	Violent self-sacrificial
Patrik Crusius	Manifesto of the El Paso attack, 2019	5 pages	Violent self-sacrificial
Adolf Hitler	*Mein Kampf*, 1925	720 pages	Violent self-sacrificial
Sayyid Qutb	*Milestones*, 1964	160 pages	Violent self-sacrificial
Karl Marx and Friedrich Engels	*Manifesto of the Communist Party*, 1848	30 pages	Ideologically extreme
Yusuf al-Qaradawi	*The Lawful and Prohibited in Islam*, 1960	177 pages	Ideologically extreme
Fjordman	*Defeating Eurabia*, 2008	343 pages	Ideologically extreme
Simone de Beauvoir	*The Second Sex*, 1949	978 pages	Moderate
Martin Luther King Jr.	*I Have a Dream*, 1963	3 pages	Moderate
Greta Thunberg	*Our House is on Fire*, 2019	3 pages	Moderate

Narrative Coding

The manifestos were analysed with a view to detecting fusion and threat, as well as other variables of interest that might impact or complement the fusion–violence link (see Figure 1). Drawing on this theoretically grounded preliminary assessment, the following five meta-category narratives and 13 sub-category narratives (see Table 3) were used for the systematic coding of the manifesto content. Although additional narrative patterns were observed (see the later "Narrative Patterns and Linguistic Proxies" section [Tables 5 and 6]), the detailed analysis presented in this section covers only narratives that were detected in at least four of the nine assessed violent self-sacrificial manifestos.

Table 3: Narrative Classifications and Definitions

Meta-Category	**Sub-Category**	**Definition**
Group alignment	In-group identification	*In-group identification* describes an individual's sense of belonging to a defined group in social psychology. Previous studies have found that in-group identification can be reflected in the use of first-person plural pronouns such as "we", "us", and "our" (Bäck et al., 2018).

	In-group identity fusion	*In-group identity fusion* is a socio-psychological concept that describes a process where personal and group identities become functionally equivalent. This dynamic is usually characterised by the use metaphors of kinship and family relatedness when talking about the in-group, e.g. words such as “brother”, “sister”, “loyalty”, “family”, “sons”, “daughters”, “our blood”, “brotherhood”, “motherland”, and “fatherland” might be used to talk about the in-group and/or fellow group members (Whitehouse & Lanman, 2014).
Out-group entitativity	Out-group slurs	*Out-group slurs* are derogatory terms used in the context of hate speech and extremist texts (Technau, 2018). They are offensive labels used to describe an entire group of people based on their ethnicity, race, gender, religion, or sexuality (Anderson & Lepore, 2013). Well-known examples are “kike”, “kufar”, “infidel”, “fag”, “negro”, “spic”, “the Jew”, the n-word, or similar terms (Jeshion, 2013).
	Out-group demonisation	*Out-group demonisation* describes “the attribution of basic destructive qualities to the other” (Alon & Omer, 2005) or the blaming of the out-group for the personal misfortunes or the in-group (Richter et al., 2018). It usually involves the framing of an out-group as bad, hostile, or threatening to the in-group. For example, studies have explained that depictions of Jews as the “devil”, “sly conspirators”, “greedy Shylocks”, or “vengeful beneficiaries” have been used to demonise them as a dangerous out-group (Schwarz-Friesel, 2015).
	Out-group dehumanisation	*Out-group dehumanisation* “involves viewing others as less than human”, for example by describing them as or comparing them with animals (Vaes et al., 2012). Beyond the literal comparison with animals, such as “monkey”, “donkey”, and “dog”, non-human related words applied to members of an out-group, such as “creature”, “tame”, and “breed”, could also be

		indicative of out-group dehumanisation (Viki et al., 2006).
Out-group threat	Existential threat to in-group	*Existential threat to in-group* summarises the idea of the in-group being threatened with physical or symbolic collective annihilation (Hirschberger et al., 2016). This might express itself in the belief that the in-group is facing a genocide or coordinated attack: for instance, some far-right extremist groups argue that white populations are facing an existential threat because they are dying out demographically due to immigration, abortion, and violence against whites (Miller-Idriss, 2021).
	Belief in out-group conspiracy	*Belief in out-group conspiracy* denotes a functionally integrated mental system that assumes that "a group of actors collude in secret to reach malevolent goals" (Bale, 2007). A linguistic analysis of the subreddit r/conspiracy found that, compared to the control group, the conspiracy theory community made more frequent use of words related to the categories "crime", "stealing", and "law" (Klein et al., 2019).
	Belief in inevitable war	*Belief in inevitable war* involves the idea that a war of races, religions, cultures, or other opposing groups is looming above the in-group and cannot be prevented, or that a war between the in- and out-group is already under way. Inevitable war narratives are closely linked to "accelerationism", which describes the desire to trigger a looming and inevitable violent escalation of existing tensions and societal collapse (Kriner et al., 2021).
Violence-condoning norms	Justification of violence	*Justification of violence* includes rational or emotional reasonings for why resorting to violence is the best or only solution (Louis et al., 2005; Spini et al., 2008). For example, research has highlighted group norms within jihadist groups that have suggested a moral justification of terrorism and violent action via the ideas of

		pre-emptive action, self-defence, or escape from a deleterious condition that requires an immediate action (Fraise, 2017; Hafez, 2007).
	Martyrdom narrative	*Martyrdom narrative* describes the glorification of violence and terrorism by framing past or future violent action by in-group members against the out-group as heroic, selfless acts that serve a bigger purpose. For example, the language and symbolism of martyrdom might appear in the form of references to "heroic martyrs", "resistance", "self-sacrifice", or "dying in glory" (Blom, 2011; el-Husseini, 2008; Hafez, 2007).
	Violent role model	*Violent role models* may be mentioned in manifestos by invoking well-known perpetrators of genocidal violence as sources of inspiration (Richter et al., 2018). For example, authors might indicate support for previously successful terrorists by expressing identification, support, or admiration (e.g. "I admire", "I salute", "I support", naming someone "Saint", "God", etc.) (Cohen et al., 2014; Davey & Ebner, 2019).
	Hopelessness of alternative solutions	*Hopelessness of alternative solutions* summarises the perceived failure of non-violent solutions, such as political, diplomatic, or other peaceful activist means. Authors of manifestos may indicate that they have "nothing to lose" or that "democracy/politics have failed" and therefore resort to more extreme solutions (Thomas & Louis, 2014).
Violence threats	Calls to violence	*Calls to violence* cover announcements of violence and/or extreme self-sacrifice committed by the author, as well as calls that encourage the manifesto's readers to engage in violence and/or self-sacrifice against a defined out-group. Words such as "kill", "shoot", "hang", "bomb", "slaughter", or "assassinate" may be indicative, but calls to violence may also reference specific weapons, such as "sniper rifles", "ammonium nitrate", etc. (Cohen et al., 2014).

Based on the above identified meta-category narratives and sub-category narratives, each manifesto was coded to indicate whether the identified narrative categories were present or absent in each manifesto. The specific counts of occurrences were not included at this stage, as the quantitative assessment of all manifestos was performed in a subsequent NLP analysis. To enable an ICR analysis in the next step, at least one evidence sentence was collected from each coded narrative occurrence in each manifesto.

Intercoder Reliability (ICR) Test

To test the coherence of the coding framework and assess the reliability of all narrative categories, an ICR analysis was conducted. Using a set of collected evidence sentences representative of each coded narrative (265 evidence sentences overall), 100 randomly selected coded phrases were extracted from all 15 manifestos. All narrative categories were represented in the set of coded sample phrases. The sequence in which the sentences appeared was randomised and any mentions of authors and pages were removed.

During the first stage of the ICR analysis, the thesis author, as well as two independent expert coders, received the 100 anonymised, randomised sample phrases, which they needed to code based on a coding manual that outlined the coding procedure and explained the pre-selected narrative classifications.

Coder 1 was Graham Macklin, Assistant Professor at the Centre for Research on Extremism (C-REX) at the University of Oslo. Coder 2 was Jacob Davey, Research Manager specialising in far-right extremism at the ISD. The expert coders received the

anonymised Excel sheets and the coding manual and returned their coding of the 100 random sample phrases. For each sample phrase, the coders could select up to two narrative classifications from a dropdown list or select the answers "other" or "none" if they felt that none of the pre-selected narratives fitted the sample phrase. Ultimately, the percentage agreement of the independent expert coding with the thesis author's original coding was calculated.

This first expert intercoder review yielded an ICR of over 80% for most narrative sub-categories, and over 90% for most meta-categories. The results for identity fusion, calls to violence, out-group dehumanisation, out-group demonisation, existential threat, martyrdom narrative, and violent role model were particularly promising. A debrief with the expert coders was used to address areas of disagreement or misunderstanding related to the coding.

The debrief showed that, even when there were discrepancies in the coders' choice of sub-categories, there was usually alignment on the overall category. For example, the category violence-condoning norms included justification of violence, hopelessness of alternative solutions, violent role model and martyrdom narratives. In many instances, the coders' replies varied within the range of these violence-condoning norms. Likewise, there was sometimes a lack of agreement on the specific sub-category within the wider out-group threat narrative, which contained existential threat to in-group, inevitable war, and belief in out-group conspiracy. There were also a few instances where coders could not agree, such as whether the language used was demonising or dehumanising, while the category "generalisation of the out-group" was considered too unspecific to be selected.

To further assess the reliability and coherence of all categories, the same intercoder review process was then repeated with 24 independent non-expert coders. In addition to providing an independent test, this second analysis enabled comparing coding consistency levels with non-expert coders possessing limited expertise and experience with the narratives and language used by violent extremists. Based on the feedback provided by the expert coders, the category "generalisation of the out-group" was replaced with the new narrative category "out-group slurs". The final ICR results based on the calculated percent agreement of the 24 non-expert coders showed a reliability rate of over 90% for almost all narrative meta-categories and sub-categories (see Table 4). More details about the IRC analysis can be found in the Appendix.

Table 4: Intercoder Reliability (ICR) Analysis Results: Experts and Non-Experts

(Meta-) Category	***Mean Expert ICR (%)***	***Mean Non-expert ICR (%)***
In-group identification	94	93
In-group identity fusion	94	96
Calls to violence	92	100
Out-group slurs	N/A	96
Out-group demonisation	92	88
Out-group dehumanisation	94	97
Out-group threat	91	96
Violence-condoning norms	97	93

Narrative Patterns and Linguistic Proxies

The coding of the manifestos exposed a range of overarching trends in the narratives and language that run through violent self-sacrificial manifestos as opposed to those found in ideologically extreme or moderate non-violent manifestos.

Fusion and Threat

In accordance with the fusion-plus-threat model for predicting violent extremism, indicators of fusion were highly prevalent in violent self-sacrificial manifestos. Eight out of nine authors of violent self-sacrificial manifestos used kinship language when referring to other members of the in-group. For example, Breivik asked: "How many of our sisters have and will be raped by Muslims?" (Breivik, 2011, p. 1281). He repeatedly referred to fellow white Europeans as "brothers and sisters", whom he claimed had been "ravaged robbed, beaten, terrorised" (Breivik, 2011, p. 1031). Breivik also warned that race-mixing would lead "the eternal loss of your extended ethnic family" (Breivik, 2011, p. 1161) and that his readers needed to prove their "allegiance and loyalty to 'your new kinsmen', to Europe and to Christianity" (Breivik, 2011, p. 1092).

Earnest called the "true anons", the followers of the far-right online community on the /pol imageboard on 8chan, his brothers: "To the true anons out there (you know who you are). You are the product of /pol/—the product of unadulterated truth. You are my brothers and the best dudes out there" (Earnest, 2019, p. 11). Ample evidence of kinship language could also be found in the texts of Hitler and Qutb. While Hitler referred to fellow Germans as brothers, Qutb described other Muslim believers as brothers whose bonds become stronger than those of biological kinship. He wrote:

“When the relationship of the belief is established, whether there by any relationship of blood or not, the Believers become like brothers” (Qutb, 2006, p. 99). Metaphors of shared blood was another fusion indicator that was commonly used in violent self-sacrificial texts. Roof wrote that there is “good white blood” in Uruguay, Argentina, Chile, and Brazil that is “worth saving” (Roof, 2015, p. 4). Earnest directed his manifesto in part to his “brothers in blood” (Earnest, 2019, p. 6).

Perhaps the most notable result from the coding was that in-group identity fusion proxies based on the psychological kinship measures did not occur in non-violent manifestos, whether ideologically extreme or moderate in nature. Instead, fusion proxies were detected in moderate manifestos only when used in reference to the whole of humanity. For instance, Martin Luther King Jr. referred to white people as “our brothers” (King Jr., 1963, p. 3) and expressed his hope for Black boys and girls “to join hands with white boys and girls as sisters and brothers” (King Jr., 1963, p. 5). He said (p. 5): “I have a dream that one day on the red hills of Georgia, the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood”. al-Qaradawi spoke both of a “brotherhood of Islâm” and a wider “brotherhood in humanity” (al-Qaradawi, 1960, p. 184). As noted earlier, fusion itself is not a predictor of violence and only becomes so when linked to out-group threat, mediated by violence-condoning norms.

In contrast to in-group identity fusion proxies, which were restricted to violent material, forms of in-group identification could be observed across all types of manifestos, including the moderate ones. Simone de Beauvoir’s gender-based identification with women, Martin Luther King Jr.’s race-based identification with Black

people and Greta Thunberg's age-based identification with the young generations were akin to al-Qaradawi's religious identification with Muslims and Fjordman's ethno-cultural identification with native Europeans. However, identification in violent self-sacrificial manifestos was characterised by a stronger use of the first-person plural, compared to non-violent manifestos. Breivik, for instance, kept referring to his in-group, using the first-person plural: "We, Europeans, …" (Breivik, 2011, p. 1087) or "We, the free, indigenous peoples …" (Breivik, 2011, p. 1203). Moreover, the identification expressions were often coupled with narratives of in-group victimisation or glorification. Roof spoke about the superiority of his in-group with which he identified: "But of course I dont [sic] deny that we are in fact superior" (Roof, 2015, p. 2). Qutb wrote: "The function of this Divine system which is given to us – we, who are the callers to Islam – is to provide a certain style of thinking, purified from all those jahili styles and ways of thinking which are current in the world and which have poisoned our culture by depriving us of our own mind" (Qutb, 2006, p. 32).

Figure 4 shows the high degree of fusion plus threat (89%) detected in the manifestos of violent authors (classified as "self-sacrificial violent"), as well as their absence (0%) in the manifestos of non-violent ("ideologically extreme" and "moderate") authors. It also illustrates the contrasting results for identification-plus-threat markers, which were common across all types of manifestos. Of the 78% of violent self-sacrificial manifestos exhibiting identification plus threat, 100% also exhibited fusion plus threat, which is consistent with this thesis's interpretation that fusion rather than identification is indeed driving violent behaviour.

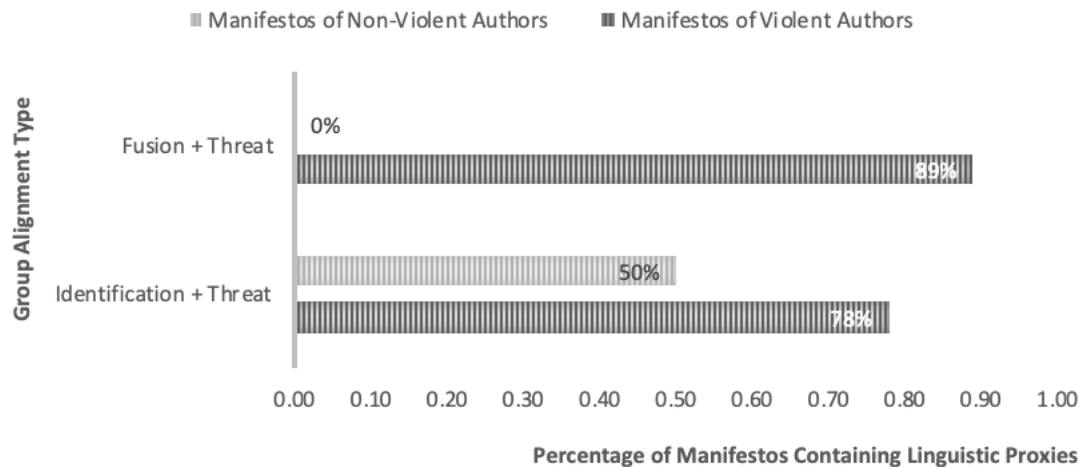


Figure 3: Group Alignment Detected in Analysed Manifestos

The idea of a threat against the in-group was a highly prevalent trait in violent self-sacrificial texts. Eight out of nine violent self-sacrificial manifestos included narratives of an existential threat against the in-group. Two in three of the ideologically extreme manifestos and one of the three moderate manifestos (Thunberg) also included an existential threat narrative. However, Thunberg's speech warned of a threat faced by the whole of humankind rather than an in-group: "I want you to act as if the house is on fire. Because it is", Thunberg (2020, p. 3) said, but she also continued, "now we all have a choice. We can create transformational action that will safeguard the future living conditions for humankind. Or we can continue with our business as usual and fail. That is up to you and me" (Thunberg, 2020, p. 2).

Words such as "exterminated", "under siege", and "replaced" were used to describe an existential threat to the in-group. Breivik (2011, p. 858) wrote "Europe is under siege by Islam" and described an existential threat to Native Europeans from a

"genocidal and evil regime committed to wiping out everything European" (Breivik, 2011, p. 1030). Earnest (2019, p. 9) warned that "we are running out of time. If this revolution doesn't happen soon, we won't have the numbers to win it." Tarrant (2019, p. 3) mentioned "an invasion on a level never seen before in history". He wrote that "we must return to replacement fertility levels, or it will kill us" (Tarrant, 2019, p. 20). Roof (2015, p. 4) painted the picture of a civil war, claiming that "white people are being murdered daily in the streets", while Baillet (2019, p. 13) spoke of "suppressed Whites". Similarly, Earnest (2019, p. 9) was afraid of "the destruction of my race", while Crusius (2019, p. 4) implied that his country was about to fall.

Hitler wrote about "the Jewish menace" and warned that "the Jew" used killings and torture "to enforce his savage fanaticism and assure domination over a great people" (Hitler, 1981, p. 217). He wrote: "If, through his Marxist faith, the Jew conquers the peoples of this world, his crown will be the death and destruction of all mankind" (Hitler, 1981, p. 45). Qutb's assessment was that "mankind today is on the brink of a precipice" (Qutb, 2006, p. 3). He called non-Muslim, secular countries "Dar-ul-Harb" (literally "the house of war") (Qutb, 2006, p. 103) and argued that "the jahili society chooses to fight and not to make peace, attacking the vanguard of Islam at its very inception" (Qutb, 2006, p. 63). It was his belief that Western regimes would try to "shake the foundations of Islamic beliefs and then gradually to demolish the structure of Muslim society" (Qutb, 2006, p. 95).

Threat narratives also came in the form of beliefs in an out-group conspiracy and inevitable war. Eight out of nine violent self-sacrificial texts contained ideas about a conspiracy of the out-group, while among the ideologically extreme it was only one out

of three, and none within the group of moderate manifestos. Example language that was indicative of conspiracy myths included “betrayal”, “sold us”, “collude against”, “conspire”, “fake”, “fraud”, and “corrupt”. Breivik (2011, p. 746) wrote that “we are slowly realising how our own elites have betrayed us and continue to do so. We are learning that they are systematically selling the European peoples into Muslim slavery.” He also claimed that “Political, cultural and media elites are colluding against the interests of free indigenous Europeans” (Breivik, 2011, p. 810). Crusius (2019, p. 1) wrote, about both Democrat and Republican politicians, that “they are either complacent or involved in one of the biggest betrayals of the American public in our history”. Hitler (1981, p. 29) wrote “Again, the hand of Fate opened my eyes to this unprecedented fraud on the people”. In *Mein Kampf*, he stated that Jews “controlled the press, the influence in art, in literature, and in the theater” (Hitler, 1981, p. 40). Qutb (2006, p. 50) described corrupted non-Muslim regimes and argued that that the earth should “be cleansed of corruption”.

Inevitable war narratives were explicitly mentioned in six out of nine violent self-sacrificial manifestos, while they did not play a role in non-violent manifestos. The idea of an inevitable war was expressed in words such as “war”, “fight”, “battle”, or “jihaad”, sometimes in combination with adjectives such as “inevitable”, “imminent”, or “looming”. Breivik (2011, p. 942) asserted that “there will be a conflict which is likely to result in enormous human suffering”. Tarrant (2019, p. 6) wrote that the conflict over the Second Amendment “will ultimately result in a civil war that will eventually balkanize the US along political, cultural and, most importantly, racial lines”. He then went on to ask: “Why should you have peace when your other brothers in Europe face certain war?”

(Tarrant, 2019, p. 43). Earnest (2019, p. 9) claimed that “civil war has just started”. Crusius (2019, p. 4) wrote that “Many people that think that the fight for America is already lost”, but claimed that “this is just the beginning of the fight for America and Europe”. For Hitler, “the fight against the enemy” was inevitable and framed as the “path to salvation for the struggling mankind of the Aryan” (Hitler, 1981, p. 435). Qutb (2006, p. 57) claimed that the struggle was “imposed upon Islam” and therefore necessary.

Other Variables of Interest

Other key patterns observed in violent authors’ language mirror the fusion–violence moderated mediators identified in previous research and visualised in Figures 1 and 2. Out-group entitativity and violence-condoning norms played a significant role in violent self-sacrificial texts. Figure 5 provides an overview of the presence of other variables of interest that were identified in the manifestos.

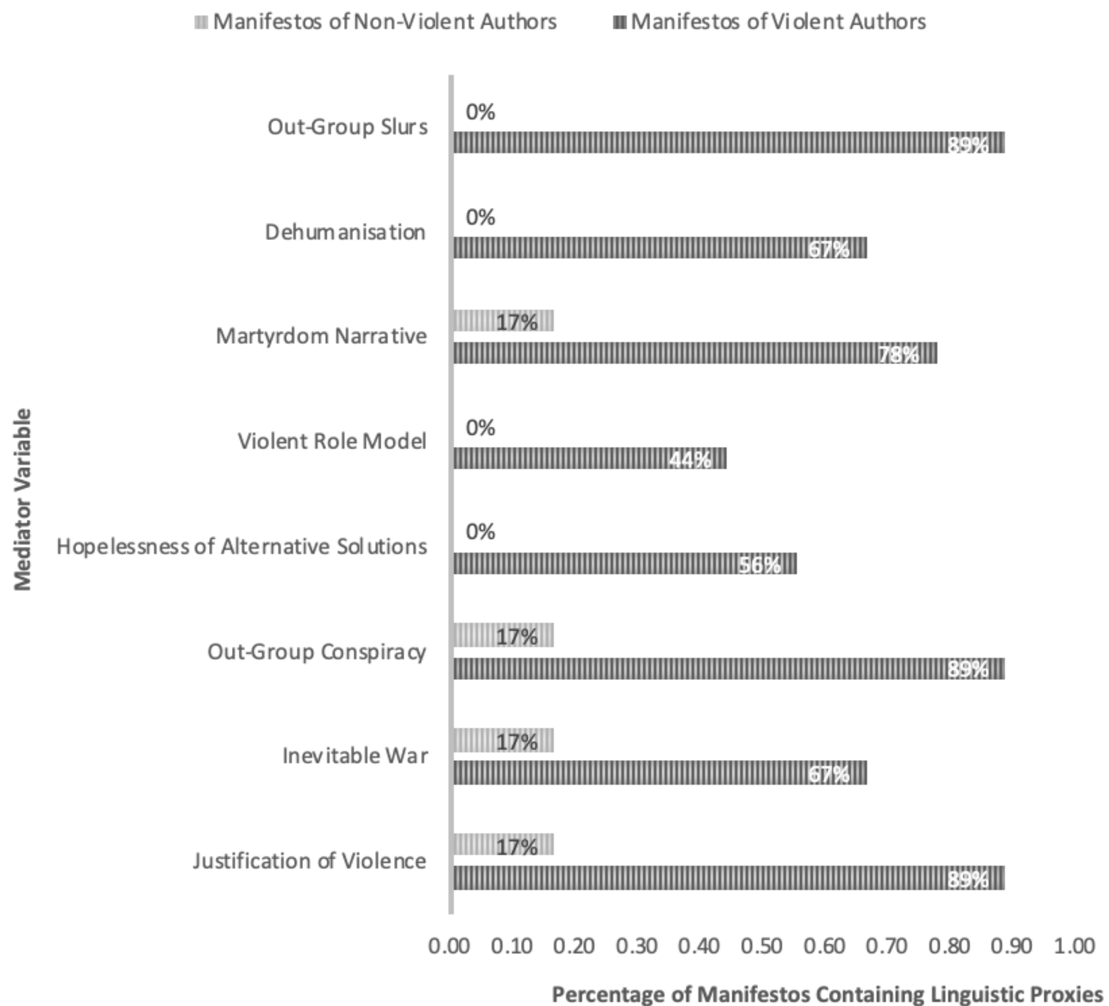


Figure 4: Other Variables of Interest Detected in the Analysed Manifestos

Out-group entitativity manifested itself linguistically in the frequent use of third-person plural pronouns, combined with the use of derogatory, demonising, or dehumanising vocabulary to denounce the out-group.

Dehumanisation was detected in six out of nine violent self-sacrificial manifestos, while it did not occur in non-violent manifestos. Words used to describe the out-group in a dehumanising fashion included "animal", "plague", "dog", "lower being", "parasite", "creature", "vermin", and "monster". Rodger (2014, p. 136) wrote "Women are like a

plague. They don't deserve to have any rights." He also called them "vicious, evil, barbaric animals" that "need to be treated as such" (Rodger, 2014, p. 136). Roof (2015, p. 2) suggested that Black people are "lower beings" and "brute animals" and likened them to dogs. Earnest (2019, p. 11) described Jews as a "squalid and parasitic race", while Hitler labelled them "slimy creatures" and compared them to "a parasite" that infests the world, a "wild beast" who tortures others, and a "spider" that sucks the blood from the people (Hitler, 1981, pp. 40, 100, 124, 125, 201).

Slurs used to describe the out-group were uniquely found in the violent self-sacrificial manifestos. For example, Roof (2014, p. 1) used offensive out-group labels such as "Niggers are stupid and violent", while Earnest (Earnest, 2019, p. 9) wrote "Spics and niggers are useful puppets for the Jew in terms of replacing Whites". Derogatory terms referring to Jews such as "kikes" were prevalent in the anti-Semitic violent self-sacrificial manifestos. For example, Baillet (2019, p. 15) wrote "The only way to win is to cut of [sic] the head of ZOG, which are the kikes". [3] Hitler (1981, p. 40) used the German equivalent of "kike" ("*Jüdlein*") as a hateful out-group slur: "If you carefully punctured this abscess with a knife, like a maggot in a rotten body who was blinded by the sudden influx of light, you would discover a Kike (*ein Jüdlein*)".

Demonisation was not only detected in all violent self-sacrificial manifestos but could also be observed in all ideologically extreme non-violent manifestos. Among the many demonising terms used for the out-group were, for instance, "traitor", "evil", "enemy", "vicious", "barbaric", "depraved", and "vile". Crusius (2019, p. 4) wrote that

[3] ZOG stands for "Zionist owned government" in reference to the anti-Semitic conspiracy myth that Jews are controlling all Western governments.

"if our country falls, it will be the fault of traitors". Breivik (2011) described the "Multicultural Marxists" as "evil" and "genocidal". Similarly, Rodger (2014, p. 136) wrote that "there is no creature more evil and depraved than the human female", while Tarrant (2019, p. 41) called Muslim minorities "the armed invader". Earnest (2019, p. 11) wrote in his manifesto: "To the Jew. Your crimes—innumerable. Your deeds—unacceptable. Your lies— everywhere." He called Jews "ugly, sinful, deceitful, cursed, and corrupt" (Earnest, 2019, p. 4) as well as "inspired by demons and Satan" (Earnest, 2019, p. 11).

Qutb (2006, p. 98) labelled all unbelievers as "the enemies of Islam". He also wrote: "Those who deviate from this system and want some other system, whether it be based on nationalism, color and race, class struggle, or similar corrupt theories, are truly enemies of mankind!" (Qutb, 2006, p. 39). Examples of out-group demonisation could also be found in Fjordman's and al-Qaradawi's ideologically extreme writing. Fjordman implied that Muslim immigrants are dangerous and part of a "barbaric culture", with a line in *Defeating Eurabia* (Fjordman, 2008, p. 249) reading: "We cannot allow the greatest nation in history to be destroyed by the planet's most barbaric cultures". Meanwhile, al-Qaradawi (1960, p. 58) warned of the "immoral behaviour and evil qualities" of the "non-believers".

In contrast with the extreme manifestos, the moderate manifestos were characterised by a low occurrence of out-group entitativity language. Martin Luther King Jr. and Simone de Beauvoir explicitly denounced the othering of an out-group in their manifestos. De Beauvoir argued that, beyond biological characteristics, the difference between man and woman does not exist and that gender is a product of society:

"Women are, among human beings, merely those who are arbitrarily designated by the word 'woman'" (de Beauvoir, 2011, p. 24); "In nature nothing is ever completely clear: the two types, male and female, are not always sharply distinguished" (de Beauvoir, 2011, p. 59). Likewise, Martin Luther King Jr. (1963, p. 5) emphasised that "that all men are created equal" and spoke of "*all* of God's children", regardless of their race or religion. He wanted Black and white people to "be able to join hands" (King Jr., 1963, p. 7). In Greta Thunberg's manifesto, climate change was framed as an urgent struggle for mankind: there is no "us" or "them" but rather a common challenge for all of us (Thunberg, 2020, p. 2).

While out-group demonisation was found, at least to some extent, across all types of manifestos, there were no instances of derogatory or dehumanising vocabulary applied to the out-group in the moderate manifestos. On the contrary, de Beauvoir, for example, criticised how women have historically been victims of demonisation and dehumanisation in literature, philosophy, and the arts. "Woman is thus doomed to Evil", she wrote, and cited Pythagoras, who said the bad principle created chaos, darkness, and the woman, as well as the Hindu Laws of Manu that defined the woman "as a vile being to be held in slavery" (de Beauvoir, 2011, p. 114).

Violence-condoning norms were identified in the form of justifications for the need to use violence, martyrdom narratives, violent role models, and the perceived hopelessness of alternative solutions. Justifications of violence played a role in eight out of nine violent self-sacrificial manifestos, as opposed to just one of the ideologically extreme manifestos (Marx and Engels), and none of the moderate manifestos. Terms such as "pre-emptive", "self-defense", "protect", "forced to fight", "need for jihaad", or

“natural struggle” were used in reference to justifications of violence. Breivik (2011, p. 776), for example, declared a “preemptive war”. He wrote an entire section about “why armed resistance against the cultural Marxist/ multiculturalist regimes of Western Europe is the only rational approach” (Breivik, 2011, p. 801). He further wrote: “We train to kill but that doesn’t mean we love violence ... We use violence only for self defence, as pre-emptive actions and as a last option” (Breivik, 2011, p. 835). Roof (2015, p. 5) used a similar line of argument in favour of violence: “I have no choice. (…) Well someone has to have the bravery to take it to the real world, and I guess that has to be me.”

Tarrant (2019, p. 13) justified the use of violent action with the alleged replacement of whites as well as jihadist attacks on non-Muslims: “I shall be forced to fight them, and hold nothing in reserve”. He wrote that he could “no longer ignore the attacks (Tarrant, 2019, p. 7). Earnest (2019, p. 2) referred to his own attack as “my act of defense”, while Crusius (2019, p. 1) wrote that “they are the instigators, not me”. He claimed to be “simply defending my country from cultural and ethnic replacement brought on by an invasion” (Crusius, 2019, p. 1). Hitler (1981, p. 427) argued that it was necessary to identify the most dangerous enemy to then “strike at him with concentrated force”. “The reasons for Jihaad” provided by Qutb included the need “to abolish all the Satanic forces and Satanic systems of life (…)” (Qutb, 2006, p. 55). He spoke about a struggle in terms of “a natural struggle between two systems which cannot co-exist for long” and his conclusion was that “Islam has no choice but to defend itself against aggression” (Qutb, 2006, p. 57).

Martyrdom narratives played a role in seven out of nine violent self-sacrificial manifestos, compared to just one of the ideologically extreme, and none in the moderate non-violent manifestos. Vocabulary used in connection with the glorification of self-sacrifice and martyrdom included, for example, “die in glory”, “sacrifice”, “knight”, “martyr”, “dying selflessly”, “immortal”, “act of preservation”, “my death”, and “defending the work of the Lord”. Breivik wrote “Equip yourself and arm up, for today you will become immortal” (Breivik, 2011, p. 1206) and “You will forever be celebrated by your people as a martyr for your country, protecting your culture and fighting for your kin and for Christendom” (Breivik, 2011, p. 947). He argued that “there is no greater glory than dying selflessly while pro-actively protecting your people from persecution and gradual demographical annihilation” (Breivik, 2011, p. 947).

Many authors framed their violence as a self-less act for society. Crusius explained that he viewed his own actions as “faultless” (Crusius, 2019, p. 4), writing “I am honored to head the fight to reclaim my country from destruction” (Crusius, 2019, p. 5). In his manifesto, Roof (2015, p. 5) wrote “Even if my life is worth less than a speck of dirt, I want to use it for the good of society”. Tarrant also argued that his “attack was a partisan action against an occupying force” and claimed to be “a lawful, uniformed combatant” (Tarrant, 2019, p. 15). He further wrote “I do not just expect to be released, but I also expect an eventual Nobel Peace prize” (Tarrant, 2019, p. 20). Earnest (2019, p. 2) announced that he was prepared to “sacrifice my future—the future of having a fulfilling job, a loving wife, and amazing kids” and preferred “dying in glory over spending the rest of his life in prison”. Both Hitler and Qutb spoke about the necessity to defend the work of God. Qutb (2006, p. 103) wrote that “the honor of martyrdom is achieved

only when one is fighting in the cause of God", while Hitler's (1981, p. 45) words were "By standing guard against the Jew, I am defending the work of the Lord".

Violent role models were mentioned by four of the nine authors of violent self-sacrificial manifestos. The authors used the names of previous attackers, often in combination with the words "inspire", "hero", or "role model". For example, Tarrant (2019, p. 19) claimed that he was influenced by the "writings of Dylan Roof and many others, but only really took true inspiration from Knight Justiciar Breivik". He also mentioned the names of the violent attackers Luca Traini, Anton Lundin Pettersson, and Darren Osbourne (Tarrant, 2019, p. 19). Earnest (2019, p. 10) stated in his manifesto that Adolf Hitler, Robert Bowers, and Brenton Tarrant had inspired him, while Crusius (2019, p. 1) wrote that he supported "the Christchurch shooter and his manifesto".

The perceived lack of alternative solutions, i.e. political or other peaceful solutions, was observed among some violent self-sacrificial manifestos. Five out of the nine analysed violent self-sacrificial texts denounced non-violent solutions as impossible or ineffective. These narratives were often expressed with words such as "democratic", "peaceful", "political", "dialogue", or "passivity", in combination with "meaningless", "failed", "end", and "vanish". For example, Breivik (2011, p. 801) wrote that the cultural conservatives of Europe have become "slaves under an oppressive, tyrannical, extreme left-wing system with absolutely no hope of reversing the damage they have caused. At least not democratically." He also stated that the "phase of dialogue has come to an end" (Breivik, 2011, p. 778), described "our weakness" as "passivity" (Breivik, 2011, p. 665), and said that "It is meaningless to participate in the democratical process" (Breivik, 2011, p. 801).

Tarrant wrote “My belief in a democratic solution vanished” (Tarrant, 2019, p. 4) and “Force is the only path to power and the only path to true victory” (Tarrant, 2019, p. 23). Crusius (2019, p. 1) warned that “America is rotting from the inside out, and peaceful means to stop this seem to be nearly impossible”. Along similar lines, Earnest (2019, p. 10) asserted that “this momentum we currently have may very well be the last chance that the European man has to spark a revolution”. Hitler made clear that he viewed political activism as ineffective: “Anyone who believes today that the question of South Tyrol can be solved by protests, speeches, local parades, etc., is either especially mischievous or a typical German middle-class citizen” (Hitler, 1981, p. 426). Qutb expressed his doubts in political solutions by saying that “all man-made individual or collective theories have proved to be failures” (Qutb, 2006, p. 4) and that “all the societies existing in the world today are jahili” (Qutb, 2006, p. 64).

Calls to Violence

As the classification of manifestos in categories from violent to non-violent was based on the authors’ actions following the publication (see above) rather than their words, the fusion–violence correlation could have been examined without an additional linguistic assessment of calls for violence and self-sacrifice. However, it may nonetheless be useful to examine linguistic indicators of violence to provide further details of the relationship between fusion language and violent language. For example, the identified lists of terms and phrases that are used to call for violence might, when independently validated and refined, be of assistance for government security services and technology companies in their efforts to develop better early violence-detection mechanisms that can help to effectively prevent attacks.

Eight out of nine of the analysed violent self-sacrificial manifestos contained explicit calls to violence against the out-group. “All category A and B traitors who continue to oppose us will be executed” Breivik (2011, p. 797) wrote. He recommended that members of his in-group must “embrace and familiarise themselves with the concept of killing women, even very attractive women” (Breivik, 2011, p. 941). “KILL ANGELA MERKEL, KILL ERDOGAN, KILL SADIQ KHAN”, Tarrant (2019, p. 40) wrote in his manifesto. Earnest (2019, p. 6) called on “every anon” reading his manifesto “to carry out attacks”. Equally, Hitler called for the use of violent force against Jews, declaring: “we need to use the most brutal weapons against a criminal group that is hostile to the state” (Hitler, 1981, p. 23). Qutb (2006, p. 49) urged all Muslims to “fight against those among the People of the Book who do not believe in God and the Last Day”.

Given the nature of the material examined, it is no surprise that many of the terrorists explicitly stated in their manifestos that they were about to carry out a violent attack. Tarrant (2019, p. 9) noted: “It was there I decided to do something, it was there I decided to take action, to commit to force. To commit to violence. To take the fight to the invaders myself.” Rodger (2014, p. 137) announced: “All of those popular people who live hedonistic lives of pleasure, I will destroy, because they never accepted me as one of them. I will kill them all and make them suffer, just as they have made me suffer.” In some instances, the violence indicators rather took the form of the announcement of self-sacrifice. Multiple terrorists suggested in their writings that they would fight and die for the in-group. “If I fail and die but kill a single jew, it was worth it,” Baillet (2019, p. 15) wrote. Earnest (2019, p. 2) stated: “I sacrifice this for the sake of my people. OUR people. I would die a thousand times over to prevent the doomed fate that the Jews

have planned for my race." "My death is likely inevitable", Crusius (2019, p. 4) declared in his manifesto.

Both the ideologically extreme and the moderate non-violent manifestos lacked indicators of violence. Additionally, many authors used language that reflected their rejection of violence. "In the process of gaining our rightful place, we must not be guilty of wrongful deeds", King Jr. (1963, p. 2) declared. He continued to plead for a "struggle on the high plane of dignity and discipline. We must not allow our creative protest to degenerate into physical violence" (King Jr., 1963, p. 2). De Beauvoir was against martyrdom, which according to her "has never changed the face of the world" (de Beauvoir, 2011, p. 184). Fjordman and al-Qaradawi also condemned the use of violence. Fjordman (2008, p. 4) wrote: "Even though I am very critical of Islam, I will not allow or support attacks on random Muslims. Those who desire this are not allowed to use my material." Al-Qaradawi was opposed to violence against human beings, unless a non-Muslim attacked first. He wrote: "Thus far we have quoted those texts which warn Muslims against killing or fighting fellow Muslims. But let no one get the impression that the life of a non-Muslim is not safe in a Muslim society, for Allâh has declared the life of every human being to be sacred, and He has safeguarded it" (al-Qaradawi, 1960, p. 179).

Tables 5 and 6 provide a detailed overview of the narratives that were detected in the manifestos, as well as their corresponding linguistic markers.

Table 5: Detailed Overview of Narratives Detected in Manifestos

Narratives	Sub-Narratives	Self-Sacrificial Violent									Ideologically Extreme			Moderate		
		Breivik	Rodger	Roof	Tarrant	Earnest	Baillet	Crusius	Hitler	Qutb	Marx and Engels	al-Qaradawi	Fjordman	de Beauvoir	King Jr.	Thunberg
Group alignment	In-group identification	Y	N[a]	Y	Y	Y	N	Y	Y	Y	Y[b]	Y	Y	Y	Y	Y
	In-group identity fusion	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N[c]	N
Violence threats	Calls to violence	Y	Y	N	Y	Y	Y	Y	Y	Y	N[d]	N	N	N	N	N
Out-group entitativity	Out-group slurs	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	N	N
	Out-group demonisation	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N
	Out-group dehumanisation	N	Y	Y	Y	Y	N	N	Y	Y	N	N	N	N	N	N
Out-group threat	Existential threat to in-group	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y[e]
	Belief in out-group conspiracy	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	N	N
	Belief in inevitable war	Y	N	N	Y	Y	N	Y	Y	Y	N	N	Y	N	N	N
Violence-condoning norms	Justification of violence	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N
	Martyrdom narrative	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	N	N
	Violent role model	Y	N	N	Y	Y	N	Y	N	N	N	N	N	N	N	N
	Hopelessness of alternative solutions	Y	N	N	Y	N	N	Y	Y	Y	N	N	N	N	N	N
Other variables	Emotional escalation	N	N	Y	Y	N	N	N	Y	N	N	N	N	N	N	N
	Individual victimhood	N	N	Y	N	N	N	Y	Y	N	N	N	N	N	N	N
	Desire for personal revenge	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N
	Male identity crises	Y	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N

Notes:

[a] The emphasis is on individual victimhood rather than the victimhood of all men.

[b] The in-group identification is in the third person, i.e. "the proletariat".

[c] The fusion proxies detected in Martin Luther King Jr.'s speech were not relating to an in-group but to the whole of humanity.

[d] There are different interpretations of Marx and Engels' calls for a revolution, which could be violent or peaceful revolution.

[e] The existential threat narrative concerns, however, not an in-group but the whole of humanity.

Table 6: Linguistic Proxies Detected in Manifestos

	Narrative	**Detected Keywords**
Group alignment	In-group identification	*“We”/”Us”/”Our” in combination with* “European”, “cultural conservative”, “Christian conservative”, “conservative”, “indigenous”, “non-Muslim”, “Justiciar Knight”, “patriot”, “martyr”, “nationalist”, “my people”, “my race”, “our race”, “anon”, “white men”, “whites”, “Aryan”, “true Muslim”, “believer”, “Muslim community”, “ummah”, “Muslim society”
	In-group identity fusion	“brother”, “sister”, “sons”, “daughters”, “kin”, solidarity”, “family”, “fellow …”, “comrades”, “my blood”, “our blood”, “bloodline”, “ancestry”, “descendant”, “ancestor”, “brethren” (these terms only indicate identity fusion when used metaphorically to describe the in-group rather than biological family)
Violence threats	Calls to violence	“executed”, “execution”, “punished”, “punishment”, “death penalty”, “kill”, “massacre”, “attack”, “destroy”, “retribution”, “revenge”, “punish”, “eradicate”, “starve”, “die”, “torture”, “behead”, “guns”, “must attack”, “must fight”, “must kill”, “give them hell”, “must play his part in this revolution”, “burn”, “shoot”, “flamethrowers”, “firearm”, “weapon”, “grenade”, “bomb”, “set fire”, “Molotov”, “fight”, “brutal steps”, “jihaad”, “bring death to”, “forcible overthrow”, “revolution”
Out-group entitativity	Out-group slurs	“kike”, “nigger”, “negro”, “spic”, “fag”, “goyim”, “golem”, “the Jew”, “global Jewry”, “pajeet”, “bitch”, “whore”
	Out-group demonisation	“traitor”, “corrupt”, “evil”, “enemy”, “our enemies”, “vicious”, “barbaric”, depraved”, “vile”, “puppets”, “perversion”, “blood libel” “crimes”, “cruel”, bloody”, genocidal”, “sinful”, deceitful”, “invader”, “poison”, parasite, “menace”, “brutal”, “ruthless”, “ “bloodsucking”, “dirty”, “deceptive”, “treacherous” “poisonous”, “oppression” , “oppressive”, “shirk” “unbeliever”, “immoral” “jahili”, “pollute”, “demolish”, “shake the foundations”, “Dar-ul-

		Harb", "arrogant", "mischievous", "criminal", "deceivers", "liars"
	Out-group dehumanisation	"animal", "plague", "impure", "brute", "dog", "lower Iq", "lower being", "inferior", "squalid" "parasitic", "parasite", "creature", "trash", "filth", "vermin", "spider", "devil", "monster", "beast", "reptile", "reptilian", "snake", "cockroach", "beneath human skin", "scum"
Out-group threat	Existential threat to in-group	"subjected to", "coerced", "brainwashed", "exterminated", "brutalised", "raped", "terrorised", "ravaged", "robbed", "replace", "subjugate", "make war upon my people" "destroyed", "overwhelmed", "under siege", "under demographical siege", "disenfranchise", "subvert", "destroy", "assault", "kill us", "kill our…", "running out of time", "last chance", "enslavement", "suffer", "economic plunder", "condemned to death", destruction of all mankind", "ill society", "at the brink of", "danger", "annihilation", "extinction", "decay"
	Belief in out-group conspiracy	"betray", "betrayal", "sell", "sold", "collude against", "colluded", "conspire", "fake", "fraud", "corruption", "corrupt", "ZOG", "Kalergi", "white genocide", "great replacement"
	Belief in inevitable war	"war", "battle", "fight", "jihaad" *in combination with* "imminent", "inevitable", "looming", "started", "already"
Violence-condoning norms	Justification for violence	"pre-emptive", "defend", "protect", "self-defense", "self-defence", "forced to fight", "no longer ignore", "act of defense", "purified", "purify", "brutal steps should have been used", "need for jihaad", "reasons for jihaad", "need for war", "the struggle is imposed upon", "natural struggle", "cannot co-exist"
	Martyrdom narrative	"die in glory", "sacrifice", "knight", "martyr", "dying selflessly", protecting our people", "immortal", "act of preservation", "my death", "defending the work of the Lord", "standing guard", "appears as the herald", "release mankind from servitude", "free from", "freed from"
	Violent role model	Mention of the names of previous terrorist attackers or violent political leaders (e.g. Breivik,

		Tarrant, Hitler, etc.) or specific attack references (e.g. Christchurch, Poway, El Paso, Utoya, Halle, etc.), *in combination with* terms that indicate perceived role model status such as “hero”, “role model”, “saint”, “inspiring”, “inspire”, “inspiration”, “support”, “influenced by”
	Hopelessness of alternative solutions	“democracy”, “democratic”, “peaceful”, political”, “system”, “politics”, “dialogue”, “passivity” *in combination with* “meaningless”, “weakness”, “failed”, “end”, “vanish”, “man-made”, “jahili”, “all societies existing”
Other variables	Emotional escalation	“hate”, “anger”, “angry”, “furious”, “hateful”, “despise”, “loathe”, “jealous”, “sad”, “tears”, “sobbing”, despair”, “shame”, “disgusted”, “outraged”, “angered”
	Individual victimhood	“victim”, “good guy”, “suffer”, “toxic”, “hell”, “mean”, “evil”, “unfair”, “injustice”, “rejected”, “insulted”, “humiliated”, “tragic life”, “cast out”, “bullied”, “weird kid”, “make fun of”, “ridicule”, “tease”, “loser”, “laughed at”, “my suffering”, “could I bear it all”, “giving up to despair”
	Desire for personal revenge	“revenge”, “to destroy”, “avenge”, “retribution”
	Male identity crises	“strong men”, weak men”, “masculinity”, “masculine”, “loser”, “rewarded by women”, “pretty girl”, “decent men”, “feminism agenda”, “feminazi”, “feminised man”, “feminisation”

Quantitative Manifesto Analysis

The NLP analysis of the manifestos was performed in R Studio based on the lists of linguistic markers identified in the qualitative text analysis. The most widely used programming languages for NLP techniques are R and Python. While both R and Python lend themselves to sophisticated word and syntax analysis, as well as visualisations of findings, Python has the additional advantage of allowing for automated content

scraping. However, the collection of manifestos analysed in this study did not require any data-mining activities from online platforms. R allowed for a quantitative examination of the content in the different manifestos based on a dictionary approach, as well as the use of its integrated statistical functions.

The R code used for this study (see the Appendix for details) reflects the effort to strike the balance between capturing the vast majority of relevant data points and limiting the number of false positives. The R code was written to capture derivations of keywords (such as nouns, as well as verbs and adjectives, or plural as well as singular forms). While this approach ensured that the number of false negatives was kept to a minimum, it also brought a higher ratio of false positives in the filtered datasets. To address this potential limitation, the filtered narrative datasets were exported from R and scanned manually for false positives. All detected false positives were subsequently removed manually and the matrix values were adjusted accordingly. False positives were defined as sentences in the filtered narrative datasets that featured one or more selected linguistic markers but did not actually reflect the definitions of these narratives (see Table 3). For example, R could not distinguish between fusion markers (kinship metaphors applied to the in-group) and mentions of real family members. Likewise, the computational results for dehumanising language directed towards the out-group also included mentions of real animals.

Computational NLP analysis on its own tends to be prone to mistakes resulting from coded forms of communication and content being taken out of context. While AI-supported NLP tools are becoming more reliable in detecting metaphors, satire, and coded hate speech and violent content, they are still characterised by high error rates

(Bansal et al., 2020; Becker et al., 2021). For this reason, the NLP-based analysis was combined with a manual review of the filtered datasets to finetune the results. To avoid biases in the manual review and maximise consistency in the coding categories, the ICR of the framework was previously tested with the help of two expert coders and 24 non-expert coders for the manual review. As described above, the ICR test yielded a score of over 90% for most narrative categories (Ebner et al., 2022a).

Manual Review

The total of over 100,000 sentences in the manifesto content was reviewed manually to remove false positives. All filtered datasets that contained less than 800 sentences were fully reviewed. In the case of large datasets with over 800 filtered sentences, a random sample of 500 sentences was selected from the filtered data for a manual review of false positives. The percentage of false positives found in these 500 sample sentences was then applied to the overall filtered data. To ensure that the reviewed sample was large enough, a confidence interval was calculated for each partial manual review:

$$CI = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

where: CI = confidence interval; $\bar{x}$ = sample mean; z= confidence level value; σ = sample standard deviation; and n= sample size.

Breivik's manifesto is a special case in point for the NLP analysis. As Kundnani pointed out in his analysis, large parts of Breivik's manifesto are a compilation of documents copied from various far-right websites and blogs (Kundnani, 2012b). To reflect only words written by the author, a manual review of cited articles and copy-

pasted passages was performed, and any passages not Breivik's writing were removed. After articles by far-right ideologues such as Fjordman, Robert Spencer, and even journalists of media outlets such as the *Daily Mail*, *Telegraph*, and *Weekly Standard* were deleted from the manifesto, a manual plagiarism check of Breivik's manifesto was conducted. This showed that he had copied entire pages from books and online resources, such as Wikipedia, often without citing them. After this adjustment, Breivik's manifesto comprised roughly 54,000 instead of 76,000 words.

Natural Language Processing (NLP) Findings

What is the shared trait of terrorist writings, when compared to non-violent manifestos? The visual matrix presented in Table 7 shows the percentage of sentences that contain relevant linguistic markers for each narrative category in each manifesto measured against the overall number of sentences in the respective manifesto. The results were adjusted to reflect the individual error margins detected via false positives in the manual review of the filtered data.

Table 7: Detected Linguistic Markers Across Manifestos (in %)

		Fusion	Existential Threat	Slurs	Demonisation	Dehumanisation	Conspiracy Belief	Inevitable War	Justification of Violence	Martyrdom Narrative	Violent Role Model	Hopelessness of Alternative Solutions	Calls to Violence
Violent	Breivik	0.3	0.4	0	0.8	0.1	0.2	0.1	0.2	0.3	0	0	1.9
	Rodger	0	0	0.2	0.6	0.5	0	0	0	0	0	0	1.7
	Roof	1.3	4	5.4	3.4	6.7	0	0	2.7	0	0	0	2
	Tarrant	6.1	5	0	8.4	0.9	1.4	0.1	0.6	0.4	0.1	0.2	6.6

	Baillet	0.6	1.3	4.4	1.3	0.6	1.3	0	0.6	0	0	0	33.3
	Earnest	2.4	2.7	8.1	5.8	1	3.7	0.3	3.7	3.1	0.7	0.7	8.8
	Crusius	2.6	3.2	0	3.8	0	2.6	1.3	2.6	1.3	1.3	1.3	7.6
	Hitler	2.5	1.1	2.2	2.7	0.6	0.2	0	0.4	0.5	0	0.1	0.2
	Qutb	0.5	0.1	0	2.5	0.6	0.1	0.1	0.7	0.3	0	0.3	0.5
Extreme	Marx and Engels	0	0.2	0	0.1	0	0	0	0.1	0	0	0	0
	al-Qaradawi	0.6	0	0	0.3	0	0	0	0	0.1	0	0	0
	Fjordman	0	0.8	0	1.2	0	1.9	0	0.1	0	0	0	0
Moderate	de Beauvoir	0	0.2	0	0.2	0	0	0	0	0	0	0	0
	King Jr.	0	1.3	0	4	0	0	0	0	0	0	0	0
	Thunberg	0	6.4	0	0	0	0	0	0	0	0	0	0

Table 7 shows that violent manifestos tended to be characterised by higher levels of identity fusion than non-violent manifestos of an ideologically extreme and moderate nature. While all other assessed variables appear to be higher in the group of violent manifestos than in the non-violent control group, our statistical tests show that only some variables were significantly different. A Mann–Whitney *U* test (Wilcoxon rank sum test) was carried out for each narrative category (significance level of $p < 0.05$), using the following formula:[4]

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

or

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

where: R_1 = sum of the ranks in sample 1; R_2 = sum of the ranks in sample 2; n_1= number of items in sample 1; and n_2= number of items in sample 2.

[4] We chose the non-parametric Mann–Whitney *U* test instead of an unpaired *t*-test due to the relatively small sample sizes.

Statistical tests were performed (see Table 8) to establish whether the difference was significant for:

A. violent versus non-violent manifestos; and

B. extreme versus non-extreme manifestos.

Table 8: Statistical Relevance of Linguistic Markers

	Fusion	Existential Threat	Slurs	Demonisation	Dehumanisation	Conspiracy Belief	Inevitable War	Justification of Violence	Martyrdom Narrative	Violent Role Model	Hopelessness of Alternative Solutions	Calls to Violence
A. Violent vs. non-violent significance ($p < 0.05$)												
Significance	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes
n_1	9	9	9	9	9	9	9	9	9	9	9	9
n_2	6	6	6	6	6	6	6	6	6	6	6	6
R_1	93.5	76.5	90	90	95.5	87	88	81	88.5	84	90	84
R_2	26.5	43.5	30	30	24.5	33	32	24	31.5	36	30	21
Critical U value	10	10	10	10	10	10	10	10	10	10	10	10
U_1	5.5	22.5	9	9	3.5	12	11	3	10.5	15	9	0
U_2	48.5	31.5	45	45	50.5	42	43	51	43.5	39	45	54
U value	5.5	22.5	9	9	3.5	12	11	3	10.5	15	9	0
B. Extreme vs. non-extreme significance ($p < 0.05$)												
Significance	No	No	No	No	No	No	No	Yes	No	No	No	No
n_1	12	12	12	12	12	12	12	12	12	12	12	12
n_2	3	3	3	3	3	3	3	3	3	3	3	3
R_1	109.5	89	105	103	109.5	108	106.5	99	106.5	102	105	94.5
R_2	10.5	31	15	17	10.5	12	13.5	6	13.5	18	15	10.5
Critical U value	4	4	4	4	4	4	4	4	4	4	4	4
U_1	4.5	25	9	11	4.5	6	7.5	0	7.5	12	9	4.5
U_2	31.5	11	27	25	31.5	30	28.5	36	28.5	24	27	31.5
U value	4.5	11	9	11	4.5	6	7.5	0	7.5	12	9	4.5

Identity fusion was found to be significantly higher in the manifestos of future perpetrators of extreme violence as opposed to the manifestos of non-violent authors, as illustrated in Table 8. Likewise, out-group dehumanisation and violence justification were found to have a strong statistical correlation with acts of violence. Other slightly less significant variables were out-group slurs, out-group demonisation, and hopelessness of alternative solutions. Belief in an existential threat to the in-group, an out-group conspiracy, as well as narratives of inevitable war, martyrdom, and violent role models were not found to be statistically higher in the manifestos of violent authors. Explicit calls to violence were statistically higher in the violent terrorist manifestos. However, due to self-censorship and the strategic circumvention of police detection among violent extremists, this variable may no longer be detectable in online groups. While violence justification was statistically more common in extreme manifestos than in non-extreme manifestos, the same cannot be said of any of the other variables.

As shown in Table 7, the manifesto of the misogynist terrorist Elliot Rodger was an exception to the observed patterns. Based on the qualitative data assessment, one potential explanation is that Rodger's world view was not based on in-group versus out-group thinking. Instead, Rodger viewed himself as being entirely alone in his fight against the rest of humanity, which fits the nihilistic "loner" mentality that is typical of the Incel (involuntary celibate) community he was loosely connected to (Hoffman et al., 2020; Sugiura, 2021). Future studies should specifically focus on the violent misogynist Incel community, which may not follow the violence risk patterns identified in other movements. In particular, mental health issues related to loneliness and their impact on

group cohesion should be further explored. Previous studies have found a high presence of depressive symptoms, suicidal tendencies, anxiety symptoms and post-traumatic stress disorder among self-identified Incels (Speckhard et al., 2021).

Limitations of the Framework

There are several important limitations with the manifesto analysis and new violence risk assessment framework. Despite efforts to select relevant comparison manifestos, including from contemporary high-profile cases, the selection is by no means exhaustive and is restricted to English-language material. Although previous studies have demonstrated the generalisability of the theoretical framework of the fusion-plus-threat model, this study only examines a small number of manifestos and is predominantly focused on far-right terrorist documents. Despite the absence of comparable pre-attack publications by jihadist and far-left perpetrators, there is a need to analyse more material, ideally with independent replications and preregistered studies, to confirm whether the patterns identified here are reliable and valid. This is a crucial step that would need to be undertaken ahead of any application in regard to designing potential interventions that rely on the content categories provided. Likewise, future control group manifestos could include a compilation of party manifestos from across the ideological spectrum, as well as less explicitly political manifestos, such as art manifestos, that might nonetheless make use of radial rhetoric.

Another limitation of the manifesto analysis is related to the differing timeframes of the selected manifestos. The earlier timeframes of the selected control group manifestos might limit their comparability with the more recent terrorist

manifestos. However, historical cases of extreme and non-extreme non-violent ideologues are more reliable data points, as it can safely be concluded that they never committed acts of violence during their lifetimes. Language in texts of the 20th century might differ from the language used in contemporary writings. For example, certain slurs were considered more socially acceptable. Nonetheless, most of the variables analysed for the purpose of this study are not expected to be impacted by time or space, as they tend to be unconscious markers of psychological processes rather than rhetorical devices.

It is important to note that the comparative qualitative content analysis is inherently subjective and relied on the assessments of the author. The identified categories were useful in providing a consistent structure for the assessments but, invariably, the researcher's degrees of freedom and subjective biases are likely to have influenced the content analysis. The reliability check conducted with experts and non-experts helped to assuage concerns about the coherence of the categories devised but did not directly validate the reliability of the full manifesto codings presented. Language tends to be subjective, contextual, and ambiguous. In the reliability checks, coders highlighted a few instances in the coding where the randomly selected phrases were long or dense enough for more than one or two narratives to be applicable. For example, "the only way to win is to cut of [sic] the head of ZOG, which are the kikes", could be read as a slur, a conspiracy theory reference, or a call to violence. Other featured sample phrases were too short or taken out of context. The phrase "My death is likely inevitable" caused many coders to choose "none" or "other" in the category drop-down list. Although within the context of the manifesto, this phrase would have been

interpreted as “martyrdom narrative”, it did not fit into any categories as an isolated phrase. Likewise, the term “jihaad” could, without further context, be interpreted as either violent or non-violent.

Furthermore, some narrative categories showed strong overlaps. For example, hopelessness of alternatives and inevitable war can be seen as inherent justifications of violence. Similarly, the line between out-group demonisation and dehumanisation was sometimes blurred and subject to interpretation. For instance, there was divergence between coders’ choices when the out-group was described as “vile anti-humans” or as “a plague”. Some coders provided the feedback that only ideological reasons provided for the use of violence should count as “justification of violence”. Pragmatic explanations of why violence would be the best approach could be seen as a wider “legitimisation of violence” but not specifically as a justification. A final point of confusion emerged when the use of the third person coincided with identity fusion markers. The phrase “The anger was never against the Jew, but always between German brothers” was identified as an example of in-group identity fusion by some coders, while others were unsure whether the manifesto author would consider themselves as part of the “German brotherhood”.

Another limitation of the research is that manifestos cannot offer comprehensive insights into their authors and are written with particular audiences in mind. As mentioned earlier, the authors may have strategically chosen their words and tailored their rhetoric to their target audiences. This does not impact the usefulness of the content for identifying linguistic cues of violence; however, it does mean that it is important to be wary of taking declarations or descriptions of motives as being accurate.

Similarly, the fact that a motive does not feature in a manifesto does not necessarily mean that it did not play a role in the author's radicalisation or that is not relevant regarding why they decided to conduct an attack. The important caveat here then is that the content and patterns identified should be approached critically. The categories identified are intended primarily to serve as helpful tools to identify common traits in the language and themes of violent extremists rather than claiming to offer a comprehensive psychological profile.

As expected, the NLP analysis provided a quantitatively more nuanced picture of recurring narrative and linguistic patterns in terrorist manifestos than the qualitative analysis. It also captured a few instances of linguistic markers that were not previously spotted in the qualitative research. For example, the qualitative manifesto analysis overestimated the role of "dehumanising" language in violent manifestos, which was statistically shown to be insignificant. While the R code used in this analysis provides a useful starting point for further research, it is important to acknowledge the need for it to be further developed and updated in a constant feedback loop with the latest vocabulary, insider references, conspiracy myths, and slurs used by extremist online communities. Moreover, it should be emphasised that this analysis is limited in its scope, as it only applied the proposed model to a limited number of online manifestos uploaded prior to terrorist attacks and compared them to non-violent political manifestos. Follow-up projects should test the framework in a diverse range of online groups, using big data to further finetune its components and linguistic markers.

R was used for the NLP analysis, coupled with manual reviews. In a next step, the model presented here can be integrated with sophisticated machine-learning language-

analysis tools such as CASM Technology's Method52. The Java-based tool Method52, which was developed jointly by the CASM at the London-based think tank Demos, combines machine learning mechanisms with NLP capabilities and enables the quantitative assessment of the specific linguistic markers and language patterns of any scraped materials, including raw or unstructured texts that are scraped from any platform (Sage Research Methods, 2016). The AI-based algorithms can be trained for specific research projects to produce sophisticated analysis of large sets of data. Demos itself used the in-house technology, for example, to measure Islamophobia in reaction to terrorist attacks and the Brexit vote in 2016 (Miller, 2016) and to assess the scale of social media misogyny (CASM, 2016b). Other Method52-based research includes a study of Islamist and tribalist online messaging during the 2017 elections in Kenya (Amanullah & Harrasy, 2017) and a report on the phenomenon of reciprocal radicalisation between online Islamist and far-right groups, whereby researchers trained the AI tool to identify different forms of inter-group hatred, in-group victimisation, and out-group demonisation narratives (Fielitz et al., 2018).

Caution in the use of linguistic markers for fusion and related variables is important due to the high sensitivity to false positives and the highly contextual nature of text-based expressions of socio-psychological drivers of violence. Manual reviews are therefore indispensable to ensure NLP-based results are accurate reflections of the relevant narrative categories that form part of the framework. False positives were particularly common in the categories "identity fusion" and "dehumanisation". This can be explained with the high rate of references to actual family and real animals rather than the metaphorical application of kinship language to the in-group or dehumanising

vocabulary to the out-group. For example, Elliot Rodger, whose manifesto contains much autobiographical information, frequently references actual family members, while al-Qaradawi extensively covers Islamic laws for family life. Likewise, al-Qaradawi explores the slaughtering of animals in relation to the teachings of Islam, while de Beauvoir's manifesto speaks in length about the biological differences between males and females in the animal world.

Conclusions from the Manifesto Analysis

Psychological profiling has a role to play in the prevention of violent extremism and terrorism, but only to the extent that it correctly identifies relevant predictors and a suitable method of detecting them. These are formidable challenges, requiring a well-substantiated theory of the motivators of terrorism and a feasible method of measuring them in noncompliant individuals. This chapter has suggested novel solutions to both problems. It provides the first evidence of the utility, validity, and reliability of a novel and theoretically informed codification schema that is anticipated to be useful for researchers and other parties examining extremist literature and written communications, including manifestos, blog posts, or chat messages.

The new method uses proxies for psychological constructs that provide a window into actors' motivational states and their capacity for murderous forms of self-sacrifice. Links between fusion and pro-group action are well-established but have previously been based on measuring fusion directly using pictorial and verbal scales, a procedure that would be impractical to adopt with uncooperative members of extremist groups. This chapter has shown, however, that those bent on carrying out atrocities

unknowingly reveal their levels of fusion by using language redolent of familial ties and psychological kinship. When this is also coupled with language expressing a strong conviction that the group is imperilled, and buttressed also by violence-condoning norms and the willingness to vilify, demonise, and dehumanise enemies, it robustly predicts a deadly mindset.

While the findings appear promising for the predictive power of the fusion-plus-threat model, they also highlight the relevance of other variables that might play a role in mediating pathways to violent extremism. This chapter therefore also supports alternative models that can be reconciled or even combined with the fusion-plus-threat model, such Sternberg's (2003, p. 303) "duplex theory of hate", Bandura's (2004, p. 121) "moral disengagement theory", Kruganski et al.'s (2018) "quest for significance model", and Louis et al.'s (2018) research on the power of group norms.

The qualitative and quantitative assessment of the 15 manifestos offers a much-needed contemporary update to earlier examinations of manifesto content, given the expanded size of the genre. Although the efforts here have been focused on showing how to recognise the language of violent extremism in groups that are already known to be violent, the proposed methods are generalisable and could be used, via early detection, to prevent at-risk individuals and groups from engaging in violence before it is too late. These methods are neither obvious nor intuitive insofar as they do not treat any of the core variables alone as predictors of violent extremism. Simply being fused with a group is not enough to motivate acts of terrorism, nor are violence-condoning norms, or even out-group hatred. Indeed, these variables in isolation may be harmless, even if their expressions may be unpleasant or offensive to many. Just like some volatile

chemicals, they only become dangerous when they are mixed together. Linguistic analysis can help in identifying when this is the case and acting more swiftly to avert threats before they are tragically realised.

Chapter 4: Investigating Violence Risk in Extreme-Right Online Groups

Introduction

Terrorist attacks from single-actor perpetrators aligned with far-right extremist online networks are on the rise (Assoudeh & Weinberg, 2021; Coolsaet & Renard, 2022). The threat of far-right extremist terrorism is often underestimated by policymakers and law enforcement when compared with jihadist violence (Ebner, 2017; Koehler, 2019). However, the 2019 Christchurch attack that killed over 50 people in two consecutive mosque attacks in New Zealand served as an international wake-up call (Bjørgo & Ravndal, 2019). Christchurch was followed by a series of copycat incidents in other geographies, including the attacks of Halle (Germany), Poway (US), El Paso (US) in 2019 and Buffalo (US) in 2022, apparently inspired by similar hatreds (Kupper et al., 2022; Ware, 2020). Meanwhile, the past decade has also been marked by the growth of internationally organised far-right terrorist networks, such as Atomwaffen Division and National Action, which established regional offshoots and aliases to continue their operations after governmental bans were imposed on them (Allen, 2019; Mattheis, 2022).

Far-right extremists, both single-actor perpetrators and organised terrorist groups, tend to endorse two narratives: (1) the so-called “great replacement” myth, which propagates the idea that the existence of white people is threatened by liberal migration policies and rising birth rates among non-white populations (Davey & Ebner, 2019); and (2) the “accelerationist” imperative, which proposes that an imminent race war needs to be accelerated by staging terror attacks (Dittrich et al., 2022). However,

this chapter argues that such ideas only pose a violence threat when combined with fusion.

The application of terrorist definitions to the far-right extremist side of the ideological spectrum has been inconsistent across different countries and regions (Koehler, 2019). The border between violent and non-violent extremist networks is often blurred, making it difficult to strike the balance between a human-rights-centred approach that upholds the right to free speech and effective early prevention systems that ensure that the plotting of violent attacks is detected and thwarted in time. To assess which groups or individuals pose a national security threat, a purely ideological assessment is unhelpful, as adherents of far-right extremist ideologies are located across the violence spectrum. For example, a representative survey conducted by Robert Pape at the University of Chicago found that roughly 16 million US citizens (75% of 21 million insurrectionists) support the far-right extremist idea of the "great replacement" (Pape, 2022a). While the same study found that around 13 million US citizens believe that the use of force is justified to restore Donald Trump's presidency, the vast majority of these respondents will likely not proceed on a violent path. Rather than focusing primarily on ideology, however, this thesis argues that it is necessary to investigate the underlying human drivers of violent extremism (Pape, 2022b).

This chapter tests the new linguistic violence risk assessment framework that was informed by the linguistic analysis of terrorist manifestos, as well as evidence-based research on identity fusion. To allow for cross-group comparison, the new text analysis framework was applied both to extremist and non-extremist groups, using the ISD's definition of extremism (see the Introduction).

Existing Literature

The nature of the link between extreme beliefs and out-group violence has been much debated in academia and in policy circles (Guhl, 2018). There has been a growing body of literature seeking to better understand modern-day far-right extremist violence, as well as online communities that have served as hotbeds for radicalisation towards anti-minority and anti-government violence (Guhl et al., 2020; Klein, 2019; Munn, 2019; Vegt et al., 2020). Bjørgo and Ravndal (2019) conceptualised extreme-right violence, noting that the scholarly focus on jihadist terrorism has led to a significant gap in research that seeks to understand patterns in far-right extremist attacks, their attackers, and their victims.

Contemporary research acknowledges the fractured far-right extremist landscape, while pointing to a growing convergence and cooperation between different movements that share common conspiracy myths, enemy perceptions, or political goals (Davey & Ebner, 2017; Stanford Center for International Security and Cooperation, 2022). Many studies have focused on the alt-right's vocabulary and subcultural codes, as well as their crossovers with other online communities, such as anime, music, and gaming subcultures (Dafaure, 2020; Fielitz & Marcks, 2019). Research on specific attacks, terrorists, and their online environments includes studies that have analysed the deadly white supremacist Charlottesville rally of 2017 (Blout & Burkart, 2020; Davey & Ebner, 2017; Peters & Besley, 2017), the Christchurch attack of 2019 (Davey & Ebner, 2019; Lakhani, 2022; Macklin, 2019), and the expansion of Atomwaffen Division (Ware, 2019). Other related publications have analysed neo-Nazi forums, such as Iron March

(Upchurch, 2021), and gaming apps hijacked by far-right extremist movements, such Discord, Steam, and Twitch (Davey, 2021).

As mentioned above, much of the existing research on far-right extremist terrorism has pointed to ideological pillars, such as the "great replacement" or "accelerationism", as sources of inspiration. An increasing number of studies have also examined manipulation and propaganda techniques used by extremist movements to lower the threshold for using physical force. In particular, the gamification of far-right terrorism has attracted much scholarly attention in recent years, as it has been capable of removing psychological inhibitions to violence by blurring the lines between fiction and reality (Ebner, 2020; Lamphere-Englund & White, 2022; Schlegel, 2021). Even though these factors may contribute to an environment in which radicalisation towards violence is likely to occur, they do not provide a robust basis for distinguishing between someone who will eventually carry out acts of extreme violence and someone who will cleave to extreme beliefs but not pose a threat to society.

In line with the thesis's overall fusion-based approach, this chapter is based on the hypothesis that in-group identity fusion is a key driver of violent extremism. Fusion alone, however, is not thought to cause radicalisation towards violence. Some of the additional risk factors identified in Chapter 3 may be expressed in beliefs associated with far-right narratives, such as the demonisation of out-groups linked to the "great replacement" myth, while "accelerationist" narratives are marked by violence-condoning norms that justify the use of physically forceful action against a pre-defined enemy group.

Approach and Methods

This chapter tests the new linguistic violence risk assessment based on the comparative manifesto analysis presented in Chapter 3. A combination of quantitative and qualitative text analysis was used to investigate the risk of violence in a range of far-right extremist online groups, as well as ideologically moderate control groups. Apart from linguistic markers of identity fusion, the framework also includes other markers that may occur in combination with fusion, such as existential threat perceptions, demonising, dehumanising, or derogatory language used against the out-group, and violence-condoning group norms.

In today's rapidly changing online environment, where extremist groups are frequently removed from, or migrate to other, platforms, the selection of far-right groups represents a convenience sample. Data were collected from different extremist groups based on three criteria: (1) access conditions that are in line with the University of Oxford's ethics framework; (2) large enough datasets of at least 1,000 messages to enable meaningful NLP analysis; and (3) the presence of extremist narratives based on an initial ethnographic assessment that drew on the above-mentioned definition of extremism. The following four extremist groups were selected for analysis:

<u>Iron March:</u> Iron March was a public web forum used by members of the neo-Nazi terrorist groups Atomwaffen Division and National Action, as well as the Scandinavian far-right extremist Nordic Resistance Movement and the Australian white supremacist Antipodean Resistance. The website was originally created in 2011 by Russian-based nationalist Alisher Mukhitdinov, who went by the name Slavros. Its users have been

linked to several terror plots and violent extremist incidents. Iron March closed in 2017 and was then turned into a searchable database by the anonymous Twitter user @jewishworker in 2019 (Singer-Emery & Bray, 2020; Upchurch, 2021).

Charlottesville 2.0: Charlottesville 2.0 was one of the key Discord servers used by the organisers of the white supremacist "Unite the Right" rally in Charlottesville, Virginia, on 11–12 August 2017. The protest brought together members of the alt-right, neo-Nazi movements, right-wing militias, and neo-Confederates. The rally escalated into several instances of violence, including the deadly car attack by James Alex Fields, which killed counter-protester Heather Heyer and injured 35 others. The server was active during the timeframe June to August 2017 before it was taken down by Discord. Its content was first leaked by Unicorn Riot (Schiano, 2017b).

Anticom: The Anticom Server was a white supremacist Discord server of the militant group Anti Communist Action. Anticom had overlapping membership with the Atomwaffen Division and shared improvised explosives and combat manuals. The Anticom Discord channel was also used by participants of the "Unite the Right" rally in 2017. The server was active during the timeframe February to September 2017. Its content was first leaked by Unicorn Riot (Schiano, 2017b).

The Right Server: The Right Server was a far-right extremist channel on Discord that was founded by former administrators of the extremist pro-Trump Centipede Central chat room. It was active during the timeframe 2017–2021. After The Right Server was deleted by Discord in 2018, a replacement channel continued under the new name The Right

Goys, which is Yiddish for non-Jews (Liao, 2018). The messages of the original server were first leaked by Unicorn Riot (Alexander, 2018).

Two easily accessible non-extremist control groups were selected for analysis to compare and contrast the results of the far-right extremist datasets. The first control group was Third Hour, a popular discussion forum for Mormons. The so-called "Forum for Latter-Day Saints" is used as a place for Mormons to discuss topics such as family, marriage and parenting, missionary work, general welfare, and conversion stories but also includes gospel boards and advice boards. There are no known links of forum members to violence or extreme interpretations of the Latter Day Saint (LDS) movement. The second control group was FootballForums.net, a popular UK discussion forum for football fans. Its users discuss British and international football, including the UEFA Champions League and Europa League, Premier League, and non-league football and domestic cups. They also have boards for off-topic discussions and more personal conversations. The forum is a mainstream football hobby forum with users who support a range of different football clubs and therefore not expected to include many radical football hooligans.

Table 9 provides an overview of the selected groups, the total message count in the relevant datasets, and the timeframes of the chat logs.

Table 9. Overview of Selected Online Groups

Group Name	Links to Violent Extremism	Cognitive Extremism Spectrum	Number of Scraped Messages	Timeframe of Scraped Messages
Iron March	Mainly violent	Ideologically extreme	1,164	Jul–Nov 2017
Charlottesville 2.0	Partly violent	Ideologically extreme	35,607	Jun–Aug 2017
The Anticom Server	Partly violent	Ideologically extreme	188,302	Feb–Sept 2017
The Right Server	Partly violent	Ideologically extreme	678,246	Sep 2017–Sep 2020
Third Hour	Non-violent	Ideologically moderate	161,977	2004–2020
FootballForums.net	Mainly non-violent	Non-ideological	304,910	2000–2020

To measure the presence of identity fusion and other relevant violence risk variables in the collected datasets, an R-based NLP analysis was performed. The R code was based on a dictionary approach that traced the linguistic markers found to be associated with higher proneness to violence in the comparative manifesto analysis of the previous chapter. To capture derivatives of relevant terms and phrases (such as different word classes, grammatical forms, and spelling mistakes) the grep R function was used. While this minimised the rate of false negatives in the analysis, a higher number of false positives had to be reviewed manually following the computational analysis.

Due to time constraints, a sampling technique was used for the manual review of large datasets. Whenever a dataset filtered for narrative-specific markers by the R code exceeded 800 messages, the author manually reviewed a random sample of 500 messages taken from the respective dataset to determine the percentage of false negatives and apply this percentage to the overall dataset. To ensure that the manually

reviewed sample was large enough and that the percentage of false positives found in the sample was representative for the entire dataset, a confidence interval was calculated. The confidence interval that was used (95% CI < ±3) means that the false positive percentages computed based on the manual sample review for larger datasets are expected to vary by a maximum margin of error of plus or minus three percentage points at a 95% probability level.

In addition to the quantitative text analysis, a qualitative assessment of the linguistic markers in the context of entire message exchanges was carried out in order to provide more in-depth insights into the socio-psychological phenomena at play in processes of online radicalisation towards violence. The pre-defined narratives and their associated linguistic markers were used as a guiding framework for the qualitative analysis.

Quantitative Results

The estimated violence risk was determined based on the Violence Risk Index, a weighted score that was developed for the purpose of this study, which reflects the statistical relevance of the different narrative categories as measured in the previously performed linguistic terrorist manifesto analysis. More specifically, the formula for the Violence Risk Index integrates the weighted average values of three categories of variables: (1) the statistically highly significant markers (weighted 0.54) fusion, out-group dehumanisation, violence justification, and explicit calls to and announcements of violence; (2) the statistically significant values (weighted 0.25) out-group slurs, out-group demonisation, and hopelessness of alternative solutions; and (3) all other relevant

variables (weighted 0.21) illustrated in Table 10 that were found to be frequently cited in terrorism literature but were not statistically significant in the analysis of terrorist manifestos.[5] The weightings were calculated based on two criteria: statistical significance; and the number of variables in the respective category (see Table 10).

Table 10. Weight Criteria for Violence Risk Index

	Fusion	**Dehumanisation**	**Justification of Violence**	**Explicit Calls to and Announcements of Violence**	**Slurs**	**Demonisation**	**Hopelessness of Alternative Solutions**	**Existential Threat**	**Conspiracy Belief**	**Inevitable War**	**Martyrdom Narrative**	**Violent Role Model**
Critical *U* value	10	10	10	10	10	10	10	10	10	10	10	10
U value	5.5	3.5	3	0	9	9	9	22.5	12	11	10.5	15
Significance	Very significant ($p < 0.5$) = *U* values < 6				Significant ($p < 0.05$) = *U* values 6–10			Not significant ($p < 0.05$) = *U* values > 10				
Equal weight distribution	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Criteria 1: Number of variables (50%)	0.33				0.25			0.42				
Criteria 2: Stat. significance (50%)	0.75				0.25			0.00				
Weight based on Criteria 1 & 2	**0.54**				**0.25**			**0.21**				

Table 11 illustrates the prevalence of the assessed narrative categories and their linguistic markers in the different datasets, as well as providing an estimated violence risk for each group based on the results of the Violence Risk Index. The percentages in the table describe the detected number of messages carrying relevant linguistic markers of each narrative category relative to the overall message count of the relevant dataset. While the percentages may seem low, they reflect the prevalence of the specific risk

[5] The following formula was used for the Violence Risk Index: Average value (stat. high variables) * 0.54 + Average value (stat. low variables) * 0.25 + Average (stat. insign. variables) * 0.21. The results were multiplied by 100 for improved readability.

markers found from across the entire dataset in each case, meaning that the comparative proportions between datasets might be considered more informative than the raw percentages.

The quantitative analysis found that the far-right extremist dataset Iron March has the highest Violence Risk Index, followed by Charlottesville 2.0, Anticom, and The Right Server. These scores mirror the groups' actual ties to violence and terrorist activities in the offline world, providing preliminary evidence that the linguistic socio-psychological assessment framework offers an accurate estimation of violence risk associated with far-right online groups. Iron March, whose members participated in terrorist and violent extremist activities, scored 88 on the violence risk scale. The groups Charlottesville 2.0 and Anticom had scores between 36 and 28, respectively, which reflects the real-world involvement of some of their members in violence at the Charlottesville rally. The Right Server scored 18 on the violence risk scale, which is in line with the group's lack of known real-world links to terrorism or extreme violence, even though a few members might have engaged in violent hate crimes. For comparison, the analysis of the non-extreme control group datasets yielded the lowest violence risk scores, with FootballForums.net scoring 4 and Third Hour scoring 1. FootballForums.net might have hosted individual users who participated in violent hooliganism, but these are likely to be isolated cases given the website's mainstream audience of football fans supporting a range of local and national teams. The popular Mormon forum Third Hour is associated with a clear non-violence stance and there are no known cases of radicalisation towards violence within this religious online community.

Table 11. Quantitative Analysis Results

	Statistically Highly Significant (Weighted 0.54)				Statistically Less Significant (Weighted 0.25)			Statistically Non-Significant Values (Weighted 0.21)					Violence Risk Value[a]	
	Fusion	Out-Group Dehumanisation	Justification of Violence	Explicit Calls for and Announcements of Violence	Out-Group Demonisation	Out-Group Slurs	Hopelessness of Alternative Solutions	Existential Threat	Belief in Conspiracy of Out-Group	Inevitable War Narrative	Martyrdom Narrative	Violent Role Model	Violence Risk Index	Estimated Violence Risk
Far-Right Extremist Groups (in %)														
Iron March	0.86	1.03	0.34	1.55	0.95	2.75	0.00	0.43	0.60	0.26	0.00	0.09	88	VERY HIGH
Charlottesville 2.0	0.42	0.12	0.07	0.25	0.35	2.47	0.01	0.10	0.12	0,01	0,03	0.00	36	HIGH
Anticom	0.11	0.19	0.03	0.25	0.21	2.10	0.00	0.09	0.05	0.01	0.02	0.00	28	HIGH
The Right Server	0.17	0.07	0.00	0.10	0.15	1.36	0.00	0.06	0.05	0.00	0.00	0.00	18	MEDIUM
Non-Extremist Control Groups (in %)														
FootballForums.net	0.14	0.00	0.00	0.03	0.14	0.04	0.00	0.00	0.00	0.00	0.00	0.00	4	LOW
Third Hour	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.03	0.00	1	LOW

Note: The following classification scheme was used for the results of the Violence Risk Index: low: < 10%; medium: 10–20%; high: 20–60%; very high: > 60%.

Qualitative Results

The aim of the qualitative text analysis was to complement the quantitative findings with deeper insights into the nature and application of the assessed violence risk variables in online conversations. The following section investigates recurring narrative patterns of fusion and other relevant markers, as well as the contexts in which they occurred across the examined datasets. This section provides actual example sentences from far-right extremist groups, which frequently include derogatory and violent language and may therefore be offensive and shocking.

In-Group Identity Fusion

As mentioned above, kinship terms applied to other members of the in-group are core to linguistic indicators of fusion. Across all far-right extremist groups, identity fusion indicators were observed, and members frequently called each other "brothers" and "sisters" in their message exchanges. A particularly high rate of fusion markers relative to the overall datasets was identified in the Iron March and Charlottesville 2.0 datasets. Members of the violent neo-Nazi forum Iron March commonly referred to each other as "brothers". The discussions ahead of the Charlottesville rally revealed anticipation of the "brotherhood" experience that group members associated with white nationalist rallies. One post read: "Anyone else here get hyped every night as the C-ville event gets closer? I'd take going to this rally ANY day over going to the Bahama's. Nothing can replace the feeling you get at a white Nationalist rally. That feeling of brotherhood and communion is so far beyond anything else that I know." Members of the Discord channels Anticom and The Right Server also made use of kinship language when talking to or about fellow group members. For example, Anticom users used phrases such as "my brother in faith"

and “Come to my defense Anticom brothers”. Likewise, The Right Server members referred to each other as “MAGA brother”, “faith race family”, and “revolutionary brothers”.

Kinship terms were often used to encourage, protect, or unify the group. For example, users in the extremist online groups encouraged each other to “help out our white brothers and sisters” and “support our conservative brothers in a defensive manner”. The idea of brotherhood and unity was emphasised by statements such as “Fuck brother wars” and “If there was ever a time where a European brotherhood would be needed, it’s right now”. One user of The Right Server wrote: “I’m an American soldier, an American. Beside my brothers and my sisters I will proudly take a stand, When liberty’s in jeopardy I will always do what’s right, I’m out here on the front lines […].” Some users also called their in-group “family” or “fam” when providing words of encouragement. For example, one user wrote “Keep it up fam. Get jacked so you can look good when you stab commies with a knife”, and another one posted “Peace be upon those who support us anon but when you’re in the real world you need protection, fam”. Instructive and encouraging words that emphasised metaphorical familiar ties included messages such as “Stay hydrated brothers and sisters”, “Just please don’t get killed brother”, “socialize my brothers”, and “HAIL VICTORY BROTHERS”.

The qualitative analysis revealed that identity fusion was often found in combination with a perceived need to defend the in-group from an existential threat posed by a demonised out-group. Expressions of kinship feelings towards the in-group were often coupled with the idea that it is necessary to defend the group from an outside threat. One user in the Charlottesville 2.0 server wrote: “He who beats down

commies with me today is my brother". Another one commented: "White brothers: battle lines are being drawn whether you like it or not, and you don't get to choose whether someone else considers you a target/enemy". Some went as far as to convey the idea of an existential threat to the in-group. Example messages read: "We are out there to imbue meaning into White cultural destruction and awaken our brothers to the reality of what is happening" or "Every little bit counts, we mustn't let our brothers and sisters be systematically executed in the land that they built!" In some instances, kinship language was used in explicit attempts to strengthen the group cohesion with messages such as "we're all brothers here" and "no more brother wars". One user said: "[…] In the end we are brothers in this fight together. We can work out Economics differences after we remove the filth."

Identity fusion also occurred in tandem with inevitable war narratives and the perceived failure of political paths to change: "I know nuclear [war] is coming, but at least we all die as brothers, right?", read one message shared in the Anticom group. Another user noted: "Civics seek to maintain a system that is bent on white genocide. We will not vote our way out of this, but I'm happy to help my white brothers & sisters any time, any place. So I'll see all you moderates in Cville." The perceived hopelessness of alternative solutions caused some users to discuss the potential use of violence in the future when the "optics" of the group would no longer be considered relevant: "I agree on both angles. We need a good spearhead that appeals to the public. But yes. If/when the time comes ... My NS brothers and I will be the sword. Waiting behind the scenes."

There was limited evidence of in-group fusion in the non-extremist datasets. A few members in the FootballForums.net control group referred to each other as

"brothers". One post used kinship language mixed with the Muslim greeting: "Asalaam-u-laikum my Brothers and Sisters. Peace be upon you all. Stay safe and happy." Members of Third Hour sometimes referred to fellow Mormons as "brothers" and "sisters" but also used kinship language when talking about people outside of the Mormon community. One user wrote: "I'm not sure if I like the idea of many people from the media bringing their sensibilities into the community of Saints. But they are our brothers and sisters after all. We'll have to accept them all."

Threat to the In-Group

Perceived external threats to the in-group included existential threat narratives, conspiracy myths about the out-group, and ideas of an inevitable war. Portrayals of an existential threat to the in-group were found across all far-right extremist datasets. "The future of the white race hangs by a nail", read a message on the Charlottesville 2.0 server. Along similar lines, users in the Anticom server commented "Patriots are getting destroyed" and "These people want to literally kill us". Conspiracy myths about the out-group were also present in all extremist groups. The anti-immigrant "great replacement" myth (often referred to as "White genocide") and the anti-Semitic "Zionist owned government" (ZOG) myth that Jewish elites control all Western governments were most widespread. Narratives of an inevitable race war were mainly observed in the Iron March dataset, reflecting the group's accelerationist ideologies. **"**Alt Right is not building a race war, the shitty economy and unemployment is doing that", posted one user in the forum.

Many messages specifically related the idea of an existential threat to the "great replacement" myth. For example, one Anticom user wrote "When races mix you destroy

the good traits of both races while maintaining the bad ones". Others commented "Allowing subversive kike elements to destroy your country" and "I'm more or less witnessing my race being crossbred to extinction right before my eyes. Makes me sick." A message on the Charlottesville 2.0 server read: "Multiculturalism is just a synonym for white genocide." Another one said: "It's not American Genocide, its White Genocide". Building on "great replacement" ideas, existential threat warnings were often used to call for action to save the white European race. Messages described Europe as being at a crossroads, where it "will literally be destroyed, or their will be massive ring-wing uprisings". Another post warned that "our Republic will balkanize if we let diversity destroy it". Threat narratives were often detected in combination with out-group "othering", as will be outlined below.

Perceived threats to the in-group were found to be rare in the non-extreme control groups. Neither existential threats nor inevitable war narratives were detected in the FootballForums.net or Third Hour datasets. Several FootballForums.net users indicated that they believed in conspiracy myths such as the idea that 9/11 was an inside job; however, only a few expressed beliefs in more extreme conspiracies. One user admitted believing that "a lot of people are 1000% lizards", including "mostly politicians, but probably a fair few celebrities too". Likewise, the use of conspiracy myths in Third Hour was rare. A few users appeared to believe that climate change was a hoax.

"Othering" of the Out-Group

Demonising, dehumanising, and derogatory language was used to describe perceived enemy groups across all far-right extremist datasets. In some instances, sentences combined different forms of out-group "othering". For example, the message "Kike

devils aren’t creative, they just promote our failed experiments” contained dehumanising, derogatory, and demonising language at once. A user in The Right Server wrote “Niggers are coded to commit crimes!”, while a message in Iron March read: “Why do [you] think that Fags aren’t a priority? They are a serious threat.” Messages that warned of “satanic thot witches”, “satanist trannies” or “satanic kikes” are other examples of multi-level “othering”.

Demonised out-groups were equated with “cancer”, “Satan”, or “witches”. Demonisation of the “globalist elites”, the media, and minority groups was particularly widespread. For example, users suggested that liberals were “trying to destroy humanity”, called globalists “cancer”, and described the Pride flag as “the symbol of the UN, of ZOG, of Silicon Valley, of the Democrat Party, of our ethnic and cultural replacement”. Others blamed “(((Globalist bankers)))” for “DESTROYING AMERICA!”[6] and viewed the “(((Media)))” as “an agent of the people who are engineering white genocide”. Examples of anti-minority demonisation included sentences such as “Mixed race people exist to destroy the society that made them possible” and “Niggers are literally destroying this country”.

The most common words used in the far-right extremist groups to dehumanise enemy groups or individuals were “filth”, “plague”, “scum”, “animals”, “monkeys”, “rabbits”, “rats”, “dogs”, “reptilians”, “snakes”, “mongrels”, and “monsters”, as well as verbs and adjectives associated with animals such as “breed”, “savage”, “feral”, and “infest”. An Iron March member described non-white people as “breeding like rabbits”,

[6] Triple brackets are an anti-Semitic symbol used by the alt-right to signal that someone or something is supposedly Jewish.

while a user on The Right Server wrote “South African monkeys are literally raping and murdering young innocent children ...”. Dehumanisation was observed to be frequently linked to calls for violence against the out-group. Referring to Black people, one user commented: “We should hunt them to extinction, like the animals they are”. Another one wrote: “Honestly, if you breed with blacks you deserve the axe”.

The most frequently used slurs across all datasets were “niggers”, “kikes”, and “faggots”, or derivatives of these terms such as “sandnigger”, “kikebook”, and “furfag”. A range of other derogatory terms were detected, such as “tranny”, “paki” “muzzie”, “spics”, “thots”, and “Japs”, among others. Example messages included: “How far will the kikes let there [sic] dogs go”, “It’s just incredible how these fags are complete emotional infants”, and “My dog hated niggers ... like a good boy”. Other users warned of “Asian streetshitters and Spics” and wrote that “Japs are degenerate scum”. “The Jew” (singular, definite article) was also classified as a frequently used derogatory anti-Semitic term due to its association with National Socialist (NS) propaganda. For instance, one user on The Right Server commented that “black people are under propaganda from the jew”. Similar to dehumanising terms, derogatory language was often found to be used in conjunction with calls for, and the glorification of, violence.

The non-extremist datasets had a visibly lower occurrence of “othering” language. There were no slurs, demonising, or dehumanising phrases detected in the Third Hour dataset. The members of FootballForums.net occasionally used derogatory words such as “fags” and “spics”. Likewise, there were only isolated cases of out-group demonisation and dehumanisation in FootballForums.net. One example of dehumanisation was the comparison of politicians and celebrities with lizards. Another

post warned of Muslims allegedly wanting to turn "Bradford, Dewsbury, and Tower Hamlets into independent states under Sharia Law" and noted "there's still a lot of us left, to outbreed yet".

Violence-Condoning Norms

Violence-condoning norms were observed in all examined far-right extremist groups from across the violence spectrum. However, they were less present in The Right Server than in the other three extremist groups. Violence endorsement took the shape of explicit calls for violence and announcements of violence, including death wishes and discussions of acquiring and using specific weapons, as well as violence justifications, martyrdom narratives, the glorification of violent role models, and the perceived hopelessness of alternative non-violent solutions.

Explicit calls to violence were found in all datasets, except those of the control groups. Posts such as "Gas the kikes. Race war now", "HANG ALL NIGGERS", and "Destroy all that does not fit in the Natural order, faggots are right near the top of that list. Rooftop faggots now" are representative examples. There were also announcements of violence by individual group members. For example, users wrote: "I wanna kill lefties", "I'd kill a communist for fun. But for a green card, I'm gonna carve him up real nice", "I want to shoot the kike", and "They are going to be massacred when this comes to an end". Sometimes, violence announcements occurred in combination with references to martyrdom narratives: "I just want to die on the battlefield so I can go to Valhalla".

A few users even posted specific plans to use violence. For example, one Anticom user shared his plans to bomb a US government building, writing: "HEY GUYS, I HAVE DIAGRAMS OF A MAJOR FEDERAL BUILDING IN THE 517 AREA CODE AND I HAVE BLUEPRINTS FOR GUNPOWDER MANUFACTURE FOR USE IN A LARGE SCALE BOMB, USED CONCURRENTLY WITH FERTILIZER". Another one detailed the steps for planning an attack with improvised explosives, from "Step 1: look at ISIS kitchen video to make a bomb out of household ingredients" to "Step 4: walk in with backpack and black block and do it Boston bomber style". Discussions about organising and using weapons, such as guns, pistols, assault rifles, and ammunition, were visible in the Iron March server, as well as the chat rooms of Charlottesville 2.0 and Anticom.

While the Iron March users were uniformly condoning violence, a stronger internal divide could be observed in the Charlottesville and Anticom servers. Some users shared thoughts such as "SHOOT ANTIFA AND THE NATIONAL GUARD", "Kill traitors and enemies", and "We are now burning actual faggots". However, others voiced their disagreement in messages such as "We really need to avoid killing them though" and "we do not bomb, we are a discord of peace". Reacting to a user who announced that he would like to "bomb a Pride Parade", another one responded: "I get that you play Roblox but stop talking about bombing pride parades and assassinations".

Death wishes and murder fantasies were common on all extremist servers. One user of Anticom wrote "I just want Jews exterminated", while a user on The Right Server noted "I'm not racist, but Asians deserve to be exterminated". The glorification of violence against out-groups was also observed among the members of all extremist groups. "Take a shot every time a nigger gets shot", read a message on The Right Server,

while a Charlottesville 2.0 user called on fellow members to "stock pile dead fags now!" Another message said: "Anytime people are shooting, bombing, or attacking Jews, a piece of my soul smiles". Justifications of violence were present in varying degrees across the far-right extremist groups. Members of Iron March justified the use of physical force against perceived enemy groups. One wrote that violence was "what the cause really needs […] Kill". Another one argued in favour of making use of the military skills acquired by some of the platform's members to "slay bodies in boot". Other users suggested that Black people or other minority groups "earned the bullet" or justified the use of violence with the need "to defend your kinfolk from invasion".

The concept of self-defence was discussed controversially in the Charlottesville 2.0 and Anticom servers. Responding to a message that called on users to "defend yourselves", one user wrote: "Defending = attacking antifa if they attack us?" Strategic conversations about the "optics" of proactive violence versus self-defence could be observed among the participants of the Charlottesville rally who were debating which weapons to bring to the rally and under what circumstances to make use of them. One user wrote: "I'll tell you how to defend yourself: if somebody punches you, do not stab them. You will go to jail." Another commented: "Not pacifists, but defenders". After the Charlottesville rally turned violent, resulting in the murder of counter-protestor Heather Heyer, as well as several other violent confrontations, the reactions in the Charlottesville 2.0 group were mixed. Some members wrote that "no one is defending violence here" and called on others to "blame those responsible for the violence and disassociate from it", while others continued to be supportive of violence against out-groups.

The non-extremist datasets did not include explicit calls for, or announcements of, violence, nor did they contain any noteworthy amounts of violence justification narratives, the glorification of violent role models, or the hopelessness of alternative non-violent solutions. Martyrdom narratives could be observed in the Third Hour discussions, mainly in reference to religious texts. In FootballForums.net, the few detected mentions of "martyr" and related words occurred in a negative context when members discussed jihadist terrorism.

Additional Observations: Announcements of Self-Harm and Group Exits

Announcements of self-harm, including suicide, were widespread across all analysed extremist groups, except Iron March. Examples of suicide wishes included "I should kill myself", "Help imma kill myself now i'm so depressed", "Scuse me, just gonna kill myself now", "Alternative to killing myself?", "I'm going to kill myself now", and "time to kill myself". Some group members asked others for help to "end their suffering". For example, users posted messages such as "Someone help me kill myself", "Me not knowing how to kill myself with a rope", and "Carve my neck and end my suffering". Other group members specifically referred to the chat conversations as causing them to become suicidal: "This chat makes me want to kill myself". One even posted an entire good-bye message into the chat, announcing their suicide at the end of the message with the words: "I am going to kill myself tomorrow night via helium. Goodbye."

A few users in the far-right extremist groups also announced their exit following violent confrontations and escalatory events such as the Charlottesville rally on 11–12 August 2017. For example, one user wrote in his self-declared "last post" that he was

scared of dying because of the group's activities and that he just wanted to live his life peacefully. "I am done with all of Politics for a long time", he concluded.[7]

Research Limitations of the Case Study

The fact that the Violence Risk Index is based on a comparative analysis of manifestos written by terrorists, as opposed to manifestos by non-violent authors, might prompt the question of whether this text analysis framework can be applied to more informal, ad-hoc forms of communication in online groups. Given that many of the socio-psychological linguistic markers used in the framework operate on an unconscious level and are not influenced by strategic wording choices, this thesis argues that they can be observed both in formal and informal forms of communication. However, future research should explore the differences in the psycholinguistic patterns of published manifestos as opposed to casual online forum conversations.

The datasets analysed in this chapter do not offer a representative account of the multitude of far-right extremist communities found in cyberspace. The aim was to cover a variety of far-right extremist groups that differ in terms of their actual links to violence rather than offer a comprehensive perspective on the phenomenon. Moreover, it is important to acknowledge that the nature of the selected groups varies in terms of

[7] Here is the complete message: "*This is my last post. After last night I am done with all of Politics for a long time. I was disapointed yes, and was really fucking scared that we were going to die. And now its gonna be a simple freedom shit and I dont really give a fuck. Politics ruined my life, it made me more stressed, tired, and angry in life, Who cares about the Jews? They are all gonna be here if you like it or not? Who cares about whites or blacks, or assad, putin nor Trump? I just want to live my life peacefully without all of this bullshit. It's horrible, no man should live like this. Call me a Cuck, Beta, Shill all you want, but I want out. This has gone the line in which I have drawn for me in politics and I do not want to go on this ride anymore, and yes, you **CAN** get off the ride, the only downside is that with the knowledge that you have, it's hard to back to Normal life. I did it before months ago and I can do it again. Overall, it was fun being with you guys, I hope we all live in peace in the ending, and I hope we get rid of the threat of Communism. See you later Space cowboys.*"

the online infrastructure and the timeframe of communication, which might impact the analysis results. Access and ethics constraints regarding the use of data from closed chatgroups that have not previously been published made it difficult to catch groups that operated in exactly the same timeframe. This means that the far-right extremist datasets date back between two and five years and can no longer be found online due to content-removal policies. The sample decisions limit the conclusions that can be drawn. There are also a number of outstanding questions, including whether there are groups that would score high on the index without being associated with real world violence, and whether there are comparable trends in jihadist groups, left-wing extremist groups, or in groups that emphasise interpersonal kinship and physical competition but are usually not violent (for instance competitive martial-arts teams). These are issues that need to be investigated in greater depth before any strong conclusions about the relationship with violence can be validated. Future studies could complement the text analysis of historic datasets with live ethnographic research in extremist online environments.

Case Study Conclusions

This chapter has described the results of a first pilot study to test the new violence risk assessment framework, which has shown that linguistic markers associated with proneness to engage in acts of extreme violence can be measured in online groups. The linguistic violence risk markers found across the datasets mirrored the known actual links of the examined groups to real-world violence and terrorist activities. This chapter was essential to develop the Violence Risk Index, which could be used as a first assessment tool by intelligence agencies, as well as tech firm's policy and intelligence

units, to provide a basis for their decisions on resource allocation. In practice, this could mean that the index would be applied alongside existing risk assessment frameworks in far-right extremism monitoring units who would then refer groups that have been identified as “high risk” to the relevant teams responsible for launching more in-depth investigations.

Despite the limitations and the need for future research outlined above, the approach adopted here promises to provide a more reliable method of addressing the needle-in-a-haystack problem posed by extremist movements whose members are mostly harmless but among whom future perpetrators of violence lurk. This thesis’s goal is to help identify at-risk online groups and individuals in the crowd before they act, using the tell-tale traces they leave behind in their writings. This chapter’s findings further strengthen the argument that the lethal cocktail of psychological factors giving rise to violent extremism is fusion and threat, combined with various forms of out-group hatred and violence-condoning norms, all of which are detectable in language used in online communications. Understanding the significance of these linguistic markers may hold the key to predicting acts of terrorism and implementing preventative measures before it is too late.

Chapter 5: Investigating Violence Risk in Online QAnon Groups

Introduction

Virtual groups that are characterised by ideological fluidity are a rapidly growing phenomenon (Office of the Director of National Intelligence, 2021). This trend, which the FBI has described as "salad bar ideologies" (U.S. Government Publishing Office, 2020), means that ideologically rooted risk assessment frameworks are largely outdated. The global conspiracy theory movement QAnon provides a case in point, which will be explored in this chapter. QAnon's confusing ideological composition, post-organisational structures, and wide-ranging membership (Enders et al., 2022) means that it does not fit into existing CT frameworks. Many national and international terrorism strategies tend to list specific jihadist, right-wing, or left-wing extremist groups, neglecting movements that transcend clear-cut ideological and organisational boundaries (Davey et al., 2021; Leidig & van Mieghem, 2021). As mentioned in the Introduction, the UN Designated Terror Groups list is almost exclusively focused on ISIS- and Al-Qaeda-related threats (United Nations Security Council, 2019). This chapter performs a risk assessment of QAnon, which has traditionally not been categorised as potential terror threat. Based on the introduced model, it explores the risk of extreme violence emanating from online communities related to QAnon.

QAnon has made headlines in recent years for its links to the storming both of the US Capitol on 6 January 2021 (Rubin et al., 2021) and the German Reichstag a few months earlier on 27 August 2020 (Keady, 2021). In both assaults on democratic institutions, the flags, banners, and T-shirts carried by the rioters featured prominent

Qanon slogans, such as WWG1WGA ("where we go one, we go all"), and QAnon symbols, such as the letter Q (Keady, 2021). According to a 2021 analysis by START, 61 QAnon adherents participated in the Capitol insurrection (Jensen & Kane, 2021). QAnon connections have also been detected in a series of violent plots and threats against political representatives in North America, Europe, and Australia (Beckett, 2020). As of September 2021, 101 QAnon followers had committed crimes in the US alone, according to START's data (Jensen & Kane, 2021). Its adherents have also inspired election-related violence, anti-vaccine protests, and pro-Russia demonstrations across the world (Brezar, 2022; Polarization and Extremism Research and Innovation Lab [PERIL] and Network Contagion Research Institute, 2020; Smirnova & Arcostanzo, 2022).

A May 2019 report by the US FBI highlighted the growing role of anti-government extremism, encompassing conspiracy theory extremists such as QAnon in criminal activity, including acts of violence (Winter, 2019). Likewise, the Norwegian Police Security Service (PST) flagged the rising threat from anti-government extremism, partly fuelled by foreign state-sponsored disinformation campaigns, in its National Threat Assessment for 2022 (PST, 2022). Nonetheless, many governments have been hesitant to label QAnon as a violent extremist threat to national security, as the advocates, followers, and sympathisers are a diverse group and do not necessarily share a distinct and clearly defined belief system. Moreover, in the past few years, policymakers and law enforcement agencies have tended to focus terrorism prevention efforts on jihadist and far-right extremist groups and networks (Amarasingam & Argentino, 2020).

QAnon began as a US-centred online subculture in 2017. On 18 October 2017, an anonymous post on 4chan's /pol (politically incorrect) board predicted that "Hilary

Clinton will be arrested between 7:45 AM – 8:30 AM EST on Monday – the morning on Oct 30, 2017”. The message was signed with “Q”, in reference to Q-level clearance in the US government. As the successor of the original “Pizzagate” conspiracy myth, QAnon continued to promote the idea that a global cabal of Satan-worshipping elites secretly controls the world and is running underground child-trafficking networks. Knitting together a variety of old and new conspiracy tropes, QAnon has become a conspiratorial master narrative that has attracted adherents from diverse ideological backgrounds based on the common denominator of anti-establishment resentment (Ebner, 2020). Today, the movement’s support base ranges from far-right extremists, hard-line conspiracy theorists, and sovereign citizens to alternative medicine esoterics, anti-vaxxers, COVID deniers, and concerned parents (Ebner, 2022). A recent study by Baker (2022) demonstrated the role of health and wellness influencers in amplifying the reach of the conspiracy myth.

In the beginning, QAnon had only several thousands of followers and was mainly focused on the US. However, less than one year after its emergence, QAnon followers were running campaigns across Europe, including efforts to boost hard-line Brexit campaigns (Ebner, 2019) and influence the discourse around the 2018 Bavarian elections (ISD, 2019a). In 2018, the ISD identified close to 30 million mentions of “QAnon” across Twitter, YouTube, and forums such as Reddit and 4chan (Ebner, 2019). By 2020, the movement had expanded to more than 4.5 million aggregate followers worldwide (Wong, 2020). Several studies have pointed to the QAnon movement’s successful exploitation of the COVID-19 pandemic and related grievances and uncertainties in the population to grow its support base (Ackerman & Peterson, 2020;

Kruglanski et al., 2020; West et al., 2021). In March 2020, ISD research registered major spikes in QAnon-related content on Twitter, Facebook, and Instagram (Gallagher et al., 2020).

Existing Literature

QAnon's sudden emergence in 2017 and swiftly rising influence both on the political arena and the security threat landscape has resulted in widespread interest both from academic and non-academic researchers across multiple disciplines. Since the movement's inception, many investigations have sought to better understand the nature, motivations, and tactics of this new online community.

Holt and Rizzuto's (2021) data-driven analysis of QAnon catchphrases on Gab, Parler, Dot-Win forums, 4chan, and 8kun during the timeframe of January 2020 to April 2021 concluded that QAnon-related slogans remained considerably higher on mainstream platforms than on alternative fringe platforms. However, the "alt-tech universe" meant that QAnon could adopt a "glocal" strategy towards disinformation, using a globally standardised and networked approach coupled with hyper-localised mobilisation channels (Ebner, 2019).

In 2021, the ISD's research highlighted the rising importance of Telegram for QAnon influencers and online groups (Gerster et al., 2021). Zihiri et al. (2022) published a study using a mixed methods approach to compare QAnon with far-right and far-left extremist communities. These researchers analysed over 3.5 million Telegram messages from these three extreme subcultures to establish QAnon's position in the wider political ecosystem of the political fringes. Their conclusion was that, despite its

mainstreamed appeal to different population segments, the QAnon community continues to share important traits with the far-right and coalesces around similar political events in the US.

Recent research and polls have highlighted the extent to which QAnon has been mainstreamed. A total of 14 congressional candidates on the ballot in the 2020 US elections openly endorsed QAnon conspiracy myths (Gallagher et al., 2020). A 2020 NPR/Ipsos poll of US adults found that 17% believed in the idea that "a group of Satan-worshipping elites who run a child sex ring are trying to control our politics and media" (NPR/Ipsos, 2020). Another representative survey conducted in the US by the Chicago Project on Security & Threat reported that there are 21 million adamant supporters of the insurrection movement with potential for violent mobilisation. According to the study, US insurrectionists are driven by two main conspiracy theories: 65% of them believe in the "great replacement" (the idea that whites are gradually being replaced by non-whites); and 54% believe in QAnon (Pape, 2021).

Many researchers have sought to understand the appeal of the QAnon cult. Matfess and Margolin (2022) established that women have played a key role in the creation and dissemination of QAnon propaganda content. According to these authors, QAnon is more gender-inclusive than traditional far-right extremist groups. As Argentino and Crawford (2021) noted, the use of female influencers and hashtags such as "#SavetheChildren" has allowed the movement to widen its reach beyond traditional audiences. Holoyda (2022) outlined the importance for forensic psychiatrists to better understand how the psychological drivers and belief systems of QAnon followers are consistent and how they are different from those held by proponents of other conspiracy

theories. Zuckerman (2019) highlighted common narrative patterns that QAnon shares with other conspiracy theories, but also argued that QAnon may be the first conspiracy that fully taps into the participatory potential of modern-day media and technology. According to a PERIL and Network Contagion Research Institute (2020) report, "QAnon bears many of the hallmarks of an augmented reality game (ARGs)".

Both in academia and in policy circles, there has been significant disagreement on the extent to which the QAnon movement poses a threat to national safety and democracy. Amarasingam and Argentino (2020) warned in 2020 that QAnon presents a novel challenge to the security forces and is a domestic terror threat in the making. Their report presented five criminal case studies with a nexus to QAnon, including one case that resulted in a terrorism charge. Likewise, Jensen and Kane (2021) conducted a study of the backgrounds of 100 QAnon sympathisers who committed crimes in the US, arguing that traditional CT strategies are unfit to address the new risk factors (e.g. mental health problems, substance use disorders, and family disruptions) found in QAnon supporters. Taking the opposite stance, Moskalenko and McCauley (2021, p. 1) argued in their research that "deradicalization efforts aimed at QAnon opinions are a waste of resources and potentially dangerous in exaggerating the QAnon threat and increasing Right-Wing perception of government over-reach". These authors combined the "two-pyramids" model of radicalisation with polling data to conclude that the threat of radical action from QAnon is "relatively small".

The underlying thesis of the two-pyramids model is that there are two types of radicalisation (radicalisation in opinion and radicalisation in action), with only the latter culminating in terrorist activity (Moskalenko & McCauley, 2021). Maskalenko and

McCauley's (2021) model bears comparison with this thesis's socio-psychologically grounded distinction between identification- versus fusion-based radicalisation. In particular, identification may be associated with extreme ideologies but not self-sacrificial behaviours (e.g. among religious fundamentalists), while fusion may drive self-sacrificial behaviour in the absence of extreme ideology (e.g. among football fans and soldiers). Nevertheless, there is insufficient evidence to conclude that QAnon radicalisation is predominantly limited to the first pyramid. This chapter seeks to contribute to this debate and help to better understand whether the prevalence of linguistic indicators can indicate underlying psychological processes that may be capable of motivating violent action among QAnon followers.

Approach and Methods

As in Chapter 4, this chapter uses a mixed methods approach, combining NLP-driven quantitative analysis with ethnographic qualitative analysis of QAnon-related communications channels. Both the quantitative and the qualitative research streams draw on the new theoretically grounded framework of violence-predicting narratives and linguistic patterns that was presented in Chapter 3. The results were compared to the outcomes in non-violent and violent control groups. The selection of variables and the metrics used to measure them was informed by the findings of the manifesto analysis in Chapter 3. Based on the findings of the preceding chapters, is argued that early detection of kinship language applied to fellow group members, in conjunction with existential threat narratives, might offer an important indicator of a higher risk for future acts of extreme violence and terrorism.

The NLP analysis was conducted in R, tracking linguistic markers that were previously identified as associated with subsequent violence. The quantitative analysis covered 200,000 messages scraped from the "Great Awakening Community", an online QAnon group of over 5,300 members exchanging messages using Discord, a VoIP and instant messaging platform originally associated with gaming communities. The "Great Awakening Community" Discord server was one of the main communication hubs for QAnon during the timeframe 2018–2019.

The results from the server were compared to content taken both from non-violent and violent control groups. The first control group was Third Hour, a popular discussion forum for Mormons with at least 30,000 users. The second control group was the white supremacist platform Iron March, which was a public web forum used by roughly 1,200 regular users, many of whom were found to have participated in terrorist activities and engaged with proscribed groups, such as National Action and Atomwaffen Division (Upchurch, 2021). A total of 1,160 messages from Iron March posted in 2017 until the forum's closure in November that year, and 160,000 messages from Third Hour in the timeframe 2004–2019 were included in the analysis. Third Hour and Iron March were selected in order to compare the QAnon content with content from two groups of online users that are characterised by a shared ideology, like QAnon, and represent the two ends of the violence spectrum for the comparative analysis.

The R code of this chapter's NLP analysis used a dictionary approach (see the Appendix for the full R code). By using the grep R function, it was possible to capture a wide range of derivations of the selected linguistic markers (e.g. nouns, as well as verbs and adjectives, both in singular and plural forms) and thereby minimise the number of

false negatives. However, this approach meant that the R-based datasets for each narrative contained a relatively high proportion of false positives, i.e. terms and phrases that were wrongly categorised as a narrative-specific linguistic marker. For instance, a common false positive that was mistaken for a fusion marker by the R code was the use of kinship language by users when speaking about their biological family rather than referring to fellow group members as "brothers" and "sisters". Likewise, the messages that the R code identified as instances of out-group dehumanisation (using linguistic markers such as "monkey", "dog", or "beast") sometimes contained references to real animals.

To address this potential limitation, the datasets of phrases captured by the R code for each narrative category were exported from R and scanned manually for false positives by the author. Based on a careful review of all messages, every detected case of a false positive was removed manually from datasets with up to 800 messages. As in Chapter 4, sample-based manual reviews of all filtered datasets were performed to eliminate false positives from the quantitative analysis results. The datasets were tested for spam activities by conducting a manual review of 100 sample messages posted by the five accounts with the highest number of messages. Based on this review, no spam accounts were detected; even the messages of the user with the highest number of messages (38,813 messages) appeared to be authentic. A general observation was that Discord is not a platform that is accommodating to bots due to its infrastructure and communication mode of live chats, which bots do not perform well in.

The quantitative analysis was complemented with real-time observations from live channels. More specifically, an ethnographic analysis was carried out across QAnon

groups that were active during the timeframe May 2020 to May 2022. In the wake of the US Capitol riots on 6 January 2021, big tech companies such as Twitter, Meta, and Google, were pressured to remove QAnon-related groups and networks from their platforms. As a result, QAnon and their adherents migrated from large social media platforms towards smaller alt-tech platforms, in particular Gab, Discord, and Telegram (Wildon & Argentino, 2021). As of July 2021, there were at least 3,500 QAnon Telegram groups and channels and more than 10,000 affiliated groups and channels in multiple languages with a global reach, according to the database of The Global Network on Extremism and Technology (GNET) (Wildon & Argentino, 2021).

The ethnographic analysis presented here covered 10 QAnon channels on Telegram and Discord, varying in their size, mode of interaction, and geographical focus. To identify relevant channels, a search for keywords that have emerged as typical QAnon phrases was performed, encompassing "QAnon", "WWG1WGA", "great awakening", "the storm", "trust the plan", "dark to light", "the military is the only way", "future proves past", "Q forces" "expose the pedos", "end the cabal", "save the children", and other phrases containing "Q" or "anon". Many of these keywords have been used in previous analysis of QAnon, such as studies conducted by the Atlantic Council's Digital Forensic Research Lab, which identified 13 QAnon phrases and terms (Holt & Rizzuto, 2021). The analysis presented here included channels operating in English, German, and French to allow for the detection of comparative differences and commonalities across different geographies.[8] Table 12 shows an overview of the selected channels.

---

[8] The narratives and associated linguistic markers used in this analysis do not vary substantially across the languages English, German, and French. For full transparency, all original quotes featured in the ethnographic observations (translated for the purpose of this thesis) are provided in separate footnotes.

Table 12: QAnon Group Selection

Channel Name	**Geography**	**Language**	**Subscribers (as of May 2022)**	**Channel Type**
Great Awakening Community	United Kingdom, United States	English	7,869	Discord, Open Posts
Q+Anons	United States, United Kingdom	English	35,164	Telegram, Central Posts
SpecialQForces	Worldwide	English	91,207	Telegram, Central Posts, Open Comments
Anons	United States	English	10,299	Telegram, Central Posts, Open Comments
Q Kingdom Family	Germany	German	7,357	Telegram, Central Posts
QFaktor Germany Die Echtzeit Analyse	Germany	German	25,201	Telegram, Central Posts
QAnon Austria	Austria	German	12,792	Telegram, Central Posts
Qlobal Change	Germany, Austria, Switzerland	German	136,943	Telegram, Central Posts
Dark to Light Channel	Germany	German, English	4,226	Telegram, Central Posts, Open Comments
QAnon Quebec/France	Canada, France	French	1,003	Telegram, Central Posts

The aim of the qualitative risk assessment was to investigate the nature and context of violence-predicting narratives and language found in QAnon groups. The author

scanned all messages published in the selected groups (see Table 1) in the timeframe May 2020 to May 2022 in order to: (1) identify occurrences of relevant linguistic markers; (2) analyse them in the context of the entire message exchange; and (3) classify them into risk categories. The risk categories "high", "medium", and "low" were used to reflect the prevalence both of fusion and threat in the messages from the assessment timeframe. The additional metric "calls to violence" was included as a comparative measure to determine to what degree the outcome of the fusion-based approach presented here aligns with traditionally used violence risk assessments via explicit threats to violence (Council of Europe, 2022; FBI, 2022; HM Government, 2015).

Quantitative Results

Table 13 summarises the relative prevalence of each of the linguistic categories in the target Great Awakening Community and the related violent and non-violent control groups. Percentages shown in the table describe the detected number of messages carrying relevant linguistic markers of each narrative category relative to the overall message count of the relevant dataset (200,555 total messages in Great Awakening, 161,977 total messages in Third Hour, and 1,164 total messages in Iron March). As in Chapter 4, the risk categories were determined based on the Violence Risk Index.[9]

[9] As in the previous chapter, the following formula was used for the Violence Risk Index: Average value (stat. high variables) * 0.54 + Average value (stat. low variables) * 0.25+ Average (stat. insign. variables) * 0.21. The results were multiplied by 100 for improved readability.

Table 13: Quantitative Analysis Results (in %)

	Statistically Highly Significant (Weighted 0.54)				Statistically Less Significant (Weighted 0.25)			Statistically Non-Significant Values (Weighted 0.21)					Violence Risk Value[a]	
	Fusion	Out-Group Dehumanisation	Justification of Violence	Explicit Calls for and Announcements of Violence	Out-Group Demonisation	Out-Group Slurs	Hopelessness of Alternative Solutions	Existential Threat	Belief in Conspiracy of Out-Group	Inevitable War Narrative	Martyrdom Narrative	Violent Role Model	Violence Risk Index	Estimated Violence Risk
The Great Awakening	0.27	0.24	0.38	0.23	0.79	0.04	0.00	0.08	0.48	0.01	0.01	0.00	24	HIGH
Control group 1: Iron March	0.86	1.03	0.34	1.55	0.95	2.75	0.00	0.43	0.60	0.26	0.00	0.09	88	VERY HIGH
Control group 2: Third Hour	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.03	0.00	1	LOW

Note: [a] The same classification scheme was used for the results of the Violence Risk Index as in Chapter 4: low: < 10%; medium: 10–20%; high: 20–60%, very high: > 60.

The analysis indicates elevated levels of fusion markers and other relevant violence risk variables, such as violence-condoning norms (including justification of violence and calls to violence), as well as "othering" language (including demonising, dehumanising, and derogatory language applied to the out-group), were found to be more common in the Great Awakening Community than in the Third Hour content, but less common than in the Iron March group.

The QAnon group's Violence Risk Index was 24 and therefore visibly higher than that of the non-violent control group Third Hour (Violence Risk Index: 1). One pattern in

the Third Hour content was that kinship language was not applied to the in-group but to all of humanity. For example, one message said: “We are told that we are all children of God. We’re all brothers and sisters.” Likewise, in Third Hour content, existential threat narratives hardly ever highlighted threats to the in-group of fellow Mormons but most often dealt with threats faced either on an individual level or by mankind.

The patterns found in the Great Awakening Community analysis demonstrate a much greater presence of anticipated predictors of violence than in average discussion forums. Notably, however, compared to the violent control group Iron March, which scored 88 on the Violence Risk Index, the levels were lower. As expected, the conversations on the terrorist Iron March website were marked by very high levels of fusion and other violence risk markers.

Qualitative Results

All QAnon groups that were examined in the ethnographic research showed some degree of identity fusion and existential threat markers. However, fusion and threat were more pronounced in some QAnon groups than in others. Four out of the 10 examined channels exhibited very high or high levels, while four were marked by medium levels and two by low levels of fusion and threat. Higher fusion and threat levels generally correlated with more instances of calls to violence, but there were a few cases that did not follow this pattern. Contrary to common practice among CT professionals, this thesis argues that it can be misleading to focus predominantly on calls to violence to determine the violent potential of any given group or individual. The fact that calls to violence are low does not necessarily mean that the risk of violence is low.

On a group level, it would be expected that high levels of calls to violence tend to go hand in hand with actual proneness to violence. They can be indicative of a group-based acceptance of violence and responsible for a violence-inciting atmosphere in the group. However, on an individual level, they are not a robust predictor for which users are most likely to engage in extreme pro-group violence. Those who use the most violence-threatening language are not necessarily identical to those that are most willing to put their lives on the line for the group.

Table 14 provides an overview of the observations made in the ethnographic research. The risk categories were determined based on the combined value of markers for fusion and threat. The following classification scheme was used to describe the prevalence of fusion, threat, and violence metrics: high: > 3 examples per 100 messages; medium: > 1 examples per 100 messages; and low: 0–1 examples per 100 messages.

Table 14: Qualitative Analysis Results

Channel Name	**In-Group Identity Fusion**	**Threat from Demonised or Dehumanised Out-Group**	**Violence-Condoning Norms**	**Estimated Violence Risk**
Great Awakening Community	High	Medium	High	High
Q+Anons	High	Medium	High	High
SpecialQForces	High	High	High	Very high
Anons	Low	Low	Low	Low
Q Kingdom Family	High	Low	Low	Medium
QFaktor Germany Die Echtzeit Analyse	Medium	Medium	Medium	Medium
QAnon Austria	Medium	High	Medium	High

Qlobal Change	Medium	Low	Low	Medium
Dark to Light Channel	Low	Low	Low	Low
QAnon Quebec/France	Medium	Medium	Low	Medium

In-Group Fusion Plus Out-Group Threat

A significant proportion of QAnon groups (40% in the selection of groups) showed very high or high levels of the linguistic hallmarks found in terrorist manifestos. Based on the levels of identity fusion and other associated risk factors found in the different geographies, English-speaking groups tended to have higher estimated violence risk levels than German and French ones. However, it is important to acknowledge that the geographic variations in threat detected in the analysis might be coincidental and should be further explored in follow-up studies.

Identity fusion markers, in combination with existential threat narratives of a demonised and/or dehumanised out-group, were highly prevalent in the channels Special Q Forces, Q+Anons, Great Awakening Community, and QAnon Austria, but could also be observed in other channels of the conspiracy theory movement. QAnon adherents frequently referred to each other as "brothers and sisters of the Awakening", "Anon Brothers and Sisters", and "Q family". One member of the Great Awakening Community wrote: "I love you all, fellow Patriots, fellow Guardians of Q ... We happy few, we band of brothers ... Never ever ever ever ever ever ever give up." Another one commented: "Stay strong my brothers and sisters. Satan wants to keep your [sic] from God and His word [...]." One user in the channel Q-Kingdom Family wrote: "I love you all,

siblings of light […] We are almost there, dear brothers and sisters […]."[10] A third one said he was deeply touched and grateful whenever he received messages from fellow members and stressed an eternal bond with his "brothers" and "sisters".[11] A message in QAnon Quebec/France read: "Rest assured my brothers and sisters, a new world is slowly taking shape and this requires the type of battle we are currently fighting".[12] One Qlobal-Change user shared a song together with the words: "How many of our brothers and sisters are playing this song before the drop?"[13] Members of the channel QFaktor Germany die Echtzeit Analyse were encouraged to trust God, the Q-Family, the plan, and the military: "Look where we are, brothers and sisters. Would you have thought that we manage to get here?", the same post continued.[14]

The emphasis on metaphorical kinship with like-minded fighters in other geographies was a common feature across different channels. **"**Support to our brothers and sisters from the Netherlands! WE WILL NOT OBEY THIS Mandatory vaccination and the fascistic covid pass ... WE ARE ALL UNITED", a message in the Anons channel said. A post by the Special Q Forces channel read: "WE LOVE OUR RUSSIAN BROTHERS and SISTERS! WE ARE THE HUMAN FAMILY." One QAnon Austria post announced that "our

---

[10] Author's translation of the German original: "*Ich liebe euch alle, liebe Lichtgeschwister (…) Wir haben es bald geschafft, liebe Brüder und Schwestern*."

[11] Original message: "*Fühl mich jeden Tag zutiefst gerührt/dankbarkeit wenn ich Eure Sprachnachrichten und Nachrichten höhre und lese! Ewige Verbindung Bruder/Schwestern*".

[12] Author's translation of the French original: "*Rassurez-vous mes frères & sœurs, un nouveau monde s'installe peu à peu et cela nécessite de livrer le type de batailles qu'on mène actuellement*".

[13] Author's translation of the German original: "*Wie viele unserer Brüder und Schwestern spielen dieses Lied vor dem Drop (LZ)*?"

[14] Author's translation of the German original: "*Alles wird gut werden und wir sollten GOTT, der Q-Familie, dem Plan und dem Militär vertrauen, die harte Arbeit der White Hats schätzen und stolz auf uns sein, dass wir so weit gekommen sind! Seht, wo wir sind, Brüder und Schwestern ... Wer hätte gedacht, dass wir es bis hierher schaffen und dass wir den Wandel zum Guten miterleben werden? Was für eine ZEIT, um am LEBEN zu sein!!!*"

French brothers and sisters just arrived in Berlin”,[15] while a comment in QFaktor Germany die Echtzeit Analyse stressed that Anons “need to pay attention to our brothers in the US and abroad”.[16]

Fusion often appeared in tandem with the idea of an existential threat against the metaphorical family of QAnon: “Brothers and sisters […] we are in a race against time and a battle against sustained, relentless propaganda. But together we can turn the tide,” read a message in the Great Awakening Community. The same channel warned that “the Globalists/Communitarians/Internationalists are attacking, Whites, guns, and free speech” with the alleged end goal of “destroying the US”. One conspiracy theorist wrote: “The evil government controls the skies, the high ground, and if you are a dissident, you will be hunted down and killed unless you have a protector. […] their objective is to reduce world population by 95%.” Another Great Awakening Community member shared the fabricated idea that “electromagnetic weapons are used to torture and subjugate countless American citizens ....” Shared suffering of the Q-family was also highlighted frequently: “Your Patience and Suffering was NOT for Nothing, Brothers and Sisters!” According to previous research, the reflection by group members on shared negative transformative experiences, such as periods of suffering and traumatic incidents, is a leading cause of identity fusion (Buhrmester et al., 2014; Vázquez et al., 2019; Whitehouse et al., 2017).

---

[15] Author‘s translation of the German original: “*Unsere Französischen Brüder und Schwestern sind in Berlin angekommen*”.

[16] Author‘s translation of the German original: “*Wir brauchen die Aufmerksamkeit unserer Brüder hier in den Vereinigten Staaten und im Ausland*”.

Existential threat frequently came in the form of inevitable war narratives: “With you on this one brother,” a user in the Great Awakening Community wrote, “war is inevitable I think and it feels like it’s getting closer every day”. Another one commented: “Now my family is being called up to end this galactic war”. To the question: “Your family? Your earthly family?”, the QAnon adherent replied “No not my human family, my soul family”. Likewise, a Special Q Forces post announced “THIS IS WAR”, calling on fellow members to prepare for the fast approaching storm. The channel QFaktor Germany die Echtzeit Analyse urged fellow patriots and Anons to be strong in anticipation of “the hardest part of the looming war”.[17] Meanwhile, a member of Q+Anons wrote: “We are in a war”. The user described the war as a battle “against the cabal, news media propaganda, banking systems, tyranny authorities”. According to a member of QAnon Quebec/France, the “the war between Dark and Light is the war between Satan and God. Between the reptilian bloodline of Satan and the human bloodline of Jesus.”

QAnon channels made frequent use of demonising and dehumanising language to denounce political opponents. For example, members of the Great Awakening Community claimed that Jews “were born predators”, speaking about their “animalistic mentality”. One message from Q+Anons reads: “Pay these parasites no mind, OUR love makes us Stronger”. Celine Dion was called a “reptile” for promoting vaccines. QAnon Quebec/France spoke of the “satanic Elite, the Iluminati [sic], the Cabal” and claimed

[17] Author‘s translation of the German original: “*Seid stark Patrioten und ANONS für den härtesten Teil des KOMMENDEN KRIEGES, wenn die Hoffnung verloren scheint*”,“*Wir müssen für alle Kinder kämpfen.Sie sind die Zukunft*”, and “*Das ist der Kampf, dem wir gegenüberstehen*”.

that "13 ROYAL families are REPTILIAN hybrids who are shape-shifters posing as HUMANS". The Dark to Light channel warned of the "globalist parasites".[18]

Violence-Condoning Norms

Even if calls for violence in isolation are not a reliable proxy for violence, they can nonetheless be a helpful marker to understand the overall atmosphere in a group. A violence-condoning group setting can, if combined with high fusion and threat, be indicative of a higher violence risk level. As expected, calls for violence were generally observed to be more common in the channels that showed higher levels of fusion and threat language. The channels Special Q Forces, Q+Anons, and the Great Awakening Community contained a higher number of direct calls for violence and the use of force compared to the other examined channels. "Divided we fall, and the only way to be united outside of government is an insurrection and concentration of force. That is what we called a target rich environment". QAnon members frequently painted the picture of a situation with no viable political or peaceful alternatives. A Q+Anons member stressed that "there is no legal path". Members who choose the illegal path can hope that "a brother will back your freedom". Another one wrote: "You may have to defend what you took thru violence". One user in the Great Awakening Community commented:

> "*It is us or them. We live and they die, or we die and they live. If they live with us, they subvert the system as they have been doing for centuries. Only by removing them and their evil influence do we have any hope of survival.*"

[18] Author's translation of the German original: "*Viele von uns in der Freiheitsbewegung verstehen die Eigenschaften des globalistischen Parasiten sehr gut, und das macht es viel einfacher für uns vorherzusagen, was sie tun werden*".

There was a recurring theme of justifying violence with the need to protect the in-group. One message in the Q+Anons channel read: “If the fools think this shit is still legitimate, they deserve the death that’s coming quickly for them. Protect yourselves against the shedding. Arm up […].” Another member of the channel wrote: “I say, if you see anyone coming to get you, shoot first and ask questions later!”

Calls for violence were often directed at prominent people supportive of the COVID-19 vaccines. “Kill Gates,” one user in Q+Anons wrote. “Why isn’t anyone killing Soros?”, another user wrote. A third user replied: “That’s an excellent question!! He’s a Jew.” The user who posed the question about murdering Soros continued: “Someone should go and cut his throat”. Prince Harry and Meghan Markle were also targeted with threats to violence. “Can I shoot them now?”, a group member wrote in response to a message that read “this couple are promoting the vaxx”, linking to a video of the speech by the Duke and Duchess of Cambridge at the Global Citizen Live event in New York. Other political opponents who were at the centre of violent threats included anti-fascist movements and minority communities such as Black people and trans people: “PEDOPHILE TRANSGENDER PINK NAZIS | ANTIFAS” (…) GOOD THING: you will die”, a post in the Q Special Forces channel said. One post in Q+Anons called on fellow members: “LYNCH this Black Lowlife Parasite, if this was my kid, this ANIMAL would not see the next light of day”.

Calls for violence frequently went hand in hand with dehumanising language applied to the targeted group or individual. For example, a Q+Anons member wrote “HANG THESE MONSTERS”, while another posted “We need to kill all these sick pedo satanic dogs!” A post about Bill Gates’s vaccine promotion was met with the comments:

"This fucking rats [sic] need to burn" and "this is evil not a human". One post on the Special Q Forces channel said "just plain filth", "they are part of SATAN", and "I will definitely not object to pedophiles getting a bullet in the head! No mercy to these fuckers! Send them to the black depths of hell!" The administrator of QAnon Austria warned that "it smells like rat" and announced that Anon "fighters" should "identify the enemy within in order to defeat the most dangerous enemy".[19]

The narrative of widespread paedophilia among the global elites was used to justify violence. "Expose the Pedos. End the Cabal", one message in the Special Q Forces channel read, and was linked to an image of an execution rope with the words "get em all". The same channel warned: "ARE YOU A PEDO OR EVEN A REPTO? Expect a bullet ANYTIME." Other threatening messages in the Special Q Forces channel included "To all BASTARDS: YOUR DEATH IS NEAR", "NO WHERE TO HIDE - the last rats got in the traps", and "THE HUNT IS ON, NO MERCY". Under every message, the slogan "Military is the only way" featured, in combination pictograms of skulls, fire, and swords, as well as "WWG1WGA" and the signature of "Q".

Research Limitations of the Case Study

Due to time and capacity limitations, only a small fraction of available QAnon channels were analysed. However, the sample selection of groups varied in their sizes, communication modes, and geographies in order to achieve a broader representation of the overall QAnon community. Despite this, limiting the quantitative analysis to only

[19] Author's translation of the German original: "*Es riecht nach einer Ratte*" and "*liebe Kämpferinnen und Kämpfer*".

one dataset per group meant that conducting meaningful inferential statistical analysis was impossible in this chapter. The associations and details provided here are thus primarily descriptive in nature and need to be subject to more robust statistical tests, with future studies ideally conducting a wider quantitative assessment using multiple datasets of QAnon and control groups.

Furthermore, even though the narrative and linguistic framework was tested using an ICR analysis, language is always subjective, contextual, and ambiguous. As mentioned in previous chapters, it is therefore important to highlight that the manual sorting of the NLP-based results can be subject to potential differing interpretations. Discord is also an idiosyncratic platform that is heavily populated by gamers and trolls, whose messages cannot always be taken at face value. The linguistic markers used in this case study, however, aim to detect patterns that can reflect unconscious and socio-psychological phenomena that are not reliant on making a distinction between satirical and serious threats.

Regarding classifications, while the selected control groups were classified into "violent" and "non-violent", these classifications only reflect a general group tendency based on known cases of members engaging in acts of violence and terrorism. However, as both groups comprise many users, most of them anonymous, it is impossible to determine the exact number of violent or non-violent members. This chapter's risk of violence assessment is not primarily based on the comparative analysis of QAnon content with the violent and non-violent control groups, but rather draws on a large body of evidence of the fusion–violence link and the systematic terrorist manifesto analysis of Chapter 3, which served as the foundation for the text-based assessment

framework. Nonetheless, it is important to be clear that the approach outlined in this thesis should not be understood as reliably singling out individuals who will commit acts of violence. The intention is more restrained and is focused around building the foundations of a framework and a linguistic toolkit that can help with risk assessment based on the proposed fusion-plus-threat model, which has been, and is being, tested in many different contexts.

To avoid outliers dominating the analysis, the raw datasheets were cross-checked for recurring usernames that made repeated use of the selected linguistic markers. While the checks offered reassurance that the linguistic analysis would not be distorted by one or several outliers, there were dozens of usernames that made more frequent use of fusion language and other variables than others, potentially signalling a greater propensity for violence than the group average. The QAnon Telegram groups that were included for the qualitative assessment varied widely in their architecture and mode of communication. While some groups were highly interactive, with messages originating from many different members, other channels were dominated by the host accounts, meaning the content was primarily that of the host. This implies that, in these specific cases, the linguistic markers might say more about the group leader's proneness to violence than the broader membership. On the other hand, continued membership in such a group does seem to imply at least a tolerance for such language. Nevertheless, the estimated risk assignments presented here should be understood with this important limitation in mind.

This chapter's finding that QAnon groups vary widely in their degrees of proneness to violence points to a need for future research. It is strongly recommended

that follow-up studies focus on further exploring why some QAnon groups and individuals showed much stronger violence predictors than others. The findings also raise further questions regarding the evolution of pro-violence group dynamics: What are psycho-social factors that make anti-establishment conspiracy myth groups turn towards violence as a viable solution? What role does in-group identity fusion play in this process? What are potential measures that could be taken to prevent identity fusion or to intervene by de-fusing members?

Case Study Conclusions

This chapter supports the hypothesis that the QAnon movement poses a risk to national security, particularly in English-speaking populations. The high prevalence of identity fusion indicators, along with external threat narratives and violence-condoning group norms, as well as demonising, dehumanising, and derogatory vocabulary in several QAnon groups, are a particularly concerning warning sign that point to an increased proneness of group members to commit acts of political violence. This assessment is further supported by the higher occurrence of calls for violence detected in QAnon channels when compared with the non-violent control group.

Taken together, the findings from the three groups offer support for the fusion-plus-threat model and illustrate how the proposed narrative and linguistic framework can be employed effectively for a computational NLP analysis of large datasets when this is followed up with a manual review of representative samples. The holistic framework seeks to provide a better way of assessing the risk of violence than simply taking calls to violence at face value. Apart from the fact that cases of high risk for

violence might be missed when only measuring calls to violence, the fusion-plus-threat approach adds to our understanding of how to manage the threats posed by today's online spaces. With increasingly strict removal policies adopted by the big tech platforms, violent extremist movements have skilfully adapted their language to evade detection and the deletion of their accounts and content. This means that even the most violent groups and individuals have started to refrain from making explicit calls to violence and would therefore easily go under the radar in most conventional monitoring systems. Even if it appears that there is a correlation between explicit threats to violence and real-world violence in this case study of channels in end-to-end encrypted messaging apps, this may no longer be the case in less private online spaces where violent extremists purposefully seek to cover up their willingness to commit violence.

The findings presented in this chapter have direct implications for research and policy. They point to the need for a new definition of violent extremism. Broad categories such as "right-wing extremism" and "Salafi-jihadist extremism" are insufficient in an era of ideologically fluid movements with the potential to resort to violence. Reflecting on this, the German intelligence agency introduced a new category for the monitoring of anti-government and anti-democracy extremist groups ("*Staatsdelegitimierer*") in 2021 to include violent extremists who no longer fit into the traditionally applied framework (Bundesamt für Verfassungsschutz, 2021). One could even ask whether radicalisation towards violence should be viewed as a phenomenon entirely distinct from ideological indoctrination, driven not by group doctrines but by identity fusion. From the perspective of security services, despite the important limitations discussed above, the analytical approach outlined in this chapter could help

with resource allocation, as it can help narrow down at-risk populations. Follow-up projects could take a user-centred approach, although this would inevitably raise ethical concerns that would need to be navigated carefully. Ultimately, fusion-based explanations of extreme violence might open new doors for potential intervention approaches, such as de-fusing members of violent extremist groups (Whitehouse, 2021).

Chapter 6: Conclusions

Drawing on existing research from the fields of social psychology, anthropology, and linguistics, this doctoral project's key contribution is in the field of terrorism studies. Its aim was to make progress on one of the most pressing questions in terrorism research: What are online predictors of radicalisation towards offline violence and how can they be detected in language? Using an interdisciplinary research design, the thesis has identified and measured recurring psychological patterns in the manifestos of convicted terrorists and compared them to the patterns observed in non-violent political manifestos. A new language-based framework grounded in fusion theory was designed and tested to assess psychological risk factors that make members of virtual groups more prone to committing acts of extreme violence.

Chapters 1, 2 and 3 of this thesis presented the study's design, situated the research performed for this doctoral project within the wider body of literature and linked its approach to relevant methods in terrorism studies, anthropology, social psychology and internet studies. The chapters found a gap in the literature regarding the study of identity formation processes in online spaces and their relationship with radicalisation towards extreme forms of violence. The literature review also pointed to a lack of studies that are focused on systematically identifying psycholinguistic patterns across terrorist writings.

In Chapter 3, the thesis presented a comparative manifesto analysis, which examined 15 manifestos, ranging from terrorist manifestos published prior to high-profile attacks in 2011–2022 (e.g. Oslo and Utøya 2011, Isla Vista 2014, Charleston 2015,

Christchurch 2019, Poway 2019, Halle 2019, El Paso 2019, etc.) to ideologically extreme and moderate manifestos that were not followed by violent attacks committed by their authors. Based on a mix of qualitative text analysis and quantitative NLP, narrative patterns and related linguistic markers were distilled and statistically assessed.

The results showed that in-group identity fusion, often in combination with an external threat, as well as dehumanising, demonising, and derogatory language applied to the out-group and violence-condoning norms, were significantly higher in violent as opposed to non-violent manifestos. While the study highlights the validity of the fusion-plus-threat model, it also supports compatible alternative models, such as Sternberg's (2003, p. 303) "duplex theory of hate", Bandura's (2004, p. 121) "moral disengagement theory", Kruglanski et al.'s (2018) "quest for significance model", and Louis et al.'s (2018) research on the power of group norms.

Chapters 5 and 6 designed and tested a new violence risk assessment framework for online groups. Drawing on the comparative manifesto analysis, as well as previous evidence on risk markers of extreme violence, a Violence Risk Index was created by integrating the linguistic risk markers based on their statistical relevance. The new violence threat assessment framework was then applied to the quantitative and qualitative analysis of a range of far-right extremist online groups across the violence spectrum, as well as the ideologically fluid conspiracy myth community QAnon and non-ideological, non-violent control groups. The calculated Violence Risk Index for each online group mirrored the respective group's links to real-world violence and terrorism. The quantitative results therefore provided preliminary evidence that the assessment framework offers an accurate estimation of violence risk associated with online groups.

The findings were further supported, explained, and contextualised by the qualitative insights provided for each analysed online group.

The alternative violence risk assessment framework presented in this thesis is an attempt to address the challenges arising from the quickly evolving online extremism landscape. New removal policies introduced by tech platforms and more visible monitoring activities by security services have prompted an increasing number of violent extremists to carefully evade keyword-based detection mechanisms in their online messages. In particular, explicit threats and announcements of violence are frequently avoided. As the lines between trolling and terrorism have become increasingly blurred, it has also become more difficult to distinguish between satire and credible threats (Ebner, 2020; RAN, 2021).

Likewise, shifts away from tightly organised terrorist groups towards loose network structures and the rise of improvised attacks by lone actor terrorists signal the need for risk assessment frameworks that do not rely on group membership (Davey et al., 2021). Finally, the increasingly ideologically fluid nature of today's extremist threats means that ideologically rooted risk assessment frameworks are largely outdated (U.S. Government Publishing Office, 2020). Examining recurring socio-psychological patterns among perpetrators of extreme forms of violence, as well as their linguistic manifestations in online communications, can therefore contribute to existing violence risk assessments by adding an analytical layer that is independent of explicit threat expressions, known group membership, or specific ideological learnings.

The new theoretically grounded model presented here has the potential to bring innovation to a space that is lagging behind the latest developments in violent extremism and online radicalisation towards violence. By combining insights from cognitive anthropology, social psychology, and linguistics into an overarching linguistic framework, this thesis has aimed to contribute to a more holistic violence threat assessment framework that can help address urgent challenges in the CT field. The proposed Violence Risk Index is unique in that it directly builds on a large body of evidence from studies in diverse contexts and draws on a theoretically grounded manifesto analysis.

By linking real-world violence to linguistic markers found in online documents by the perpetrators, this doctoral project has sought to make progress regarding the question of how to bridge the online–offline gap in research on violent extremism. In online spaces, it might help tackle the problem of winnowing out from a large population of ideological extremists which ones are most likely to commit hate crimes or acts of terrorism. While many ideological extremists are highly identified with their groups, those who are also strongly fused and show other relevant violence risk markers present the most significant threat, and so the challenge is to find these individuals before it is too late.

Implications for Theories on Radicalisation, Violent Extremism, and Terrorism

This thesis reviewed a wide range of psychological research on radicalisation, violent extremism, and terrorism, arguing that it has had limited success in establishing the underlying causes of these phenomena, making it difficult to establish effective methods

of preventing or managing them. Despite recent shifts away from simplistic theories towards multi-dimensional models that recognise the importance of psychological factors, there is potential to learn from evidence-based research on the role of identity and shared history in violent extremism.

The findings of this thesis might also have implications for future theoretical understanding of radicalisation, violent extremism, and terrorism. The model developed in this thesis suggests that identity fusion drives violent extremism when it is combined with additional variables such as perceived out-group threat, violence-condoning norms, and perceived out-group entitativity. As such, it argues that extreme ideology is a product of strong forms of identification rather than fusion and cannot in itself explain violent extremism. Indeed, the presence of fusion may be what separates radicalisation pathways towards violence from those radicalisation pathways that never end in violent expressions.

It is therefore fruitless to view extreme ideology *per se* as an indicator of violent extremism and terrorism. Common underlying narrative patterns of extremist ideologies (such as the idea of a threat from a demonised and/or dehumanised out-group and violence-condoning norms) can, however, be integral parts of extremist ideologies. For example, the idea of an existential threat is a key feature of the great replacement myth propagated by the wider white nationalist movement, while violence-condoning norms are inherently present in accelerationist and defensive jihad ideas. The idea that it is imperative to protect the white race or the global Ummah from a destructive outside threat sets the scene for violent action when combined with identity fusion dynamics. Therefore, these ideological elements may only be useful as a

diagnostic tool for detecting would-be terrorists when they occur in tandem with the more central violence risk factor: identity fusion.

The combination of socio-psychological variables that this thesis has identified as a statistically significant pattern in terrorists may be a helpful starting point for future global frameworks seeking to define terrorism. As discussed in the Introduction, definitions of “terrorism” vary widely across countries and international bodies (HM Government, 2007). ”Extremism” remains an even more ambiguous label, and its inflammatory use has led to inconsistencies and interpretations motivated by self-interest. This absence of a universally accepted, international definition of the term “terrorism” in particular has been a significant obstacle to finding global solutions that can tackle modern-day international terrorist networks. It might therefore be useful to distil the core elements that terrorists share, independent of the ideological, religious, political, and cultural context in which they emerge and operate. The analysis performed for this doctoral project may contribute to the search for such underlying socio-psychological patterns found in terrorists across the ideological spectrum to inform future definitions of terrorism.

Research Limitations and Follow-Up Projects

A range of research limitations were identified in the different data-gathering and -analysis stages of this doctoral project, which points to the need for future projects. First of all, this project was marked by challenges related to data availability and access conditions for the gathering of violent and non-violent extremist group datasets. The selection of online groups analysed in this thesis is not a representative sample that

reflects the entire far-right extremism landscape in cyberspace. It is rather a convenience sample of virtual groups from across the violence spectrum that were suitable for the purpose of this analysis.

Moreover, the data collection for this thesis was challenging due to the widely differing architectures of the online platforms that needed to be scraped to obtain the relevant datasets. The choice to make use of a commercial web scraping service instead of a self-written Python scrip for each platform can be viewed as a limitation. Even though many social scientists choose to obtain data via professional web-scraping services today, some scientists, in particular in the fields of computer science and internet studies, still write their own scripts. Ultimately, the decision should be based on the research project's focus and aims, as well as time and budget considerations. Despite efficiency gains, a clear disadvantage of using professional services for the data collection is that the web-scraping codes cannot be made public or turned into a reference for future research. While the R codes that were written for this project's analysis of the different datasets can be shared with fellow researchers who might find them useful for their own projects, the web-scraping codes cannot be made available.

The new violence risk assessment framework, which seeks to overcome the online–offline duality of violent extremism, carries inevitable limitations in its design. While the framework was developed based on previous evidence concerning psychological (and linguistic) violence predictors found in offline settings, as well as a comparative language-based analysis of manifestos of authors with varying levels of real-world violence records, it was subsequently applied to the analysis of content from online groups across the violence spectrum. However, neither offline group dynamics

nor the process of manifesto writing can be fully equated with the message exchanges in online groups. There are clear differences in the communication modes, as well as the timeframe and intentionality of the texts produced by manifesto authors as opposed to members of online groups. While some manifesto authors took weeks or months to draft their documents (e.g. Breivik), carefully choosing their words with particular audiences in mind (e.g. Tarrant), online group members may post their messages, comments, and replies in a much quicker and more spontaneous fashion. These divergences in the communication modes need to be highlighted as a potential limitation to the comparability of the bodies of texts that were analysed in this thesis, even though they are not expected to significantly impact the outcomes of the variables examined in the Violence Risk Index. As mentioned earlier, the focus on unconscious linguistic markers of socio-psychological metrics means that their analysis is largely independent of strategic rhetorical choices. Furthermore, this thesis argues that language can be both the outcome and the cause of identity-transformation processes, meaning that the "chicken-or-egg" dilemma concerning which was first (unconscious identity fusion or consciously used kinship language) does not need to be resolved in this thesis.

Another limitation stems from the highly contextual nature of online messages. The manual reviews to remove false positives from the NLP analysis results were informed by the coding framework, which was previously tested in an ICR analysis with the help of two expert coders and 24 non-expert coders. However, the messages in the manual sample reviews were often ambiguous and subject to interpretation. For example, "plague" could be read as either demonisation or dehumanisation, depending

on the context. The sentence “I am fighting for what I believe is right, not dreaming of some goofy revolution” could be interpreted as a physical or a metaphorical fight. This thesis sought to address these challenges in the additional qualitative assessment that explored the nature and context of messages in more depth. Finally, it is important to emphasise that, while the Violence Risk Index can be used as an evidence-based aid to assess which groups deserve more attention from intelligence agencies than others, it cannot replace human judgement and will need to be combined with manual investigations.

More research is needed to refine the linguistic risk assessment framework and to test the Violence Risk Index using different types of datasets. In particular, Chapters 4 and 5 were explorative in nature and sought to lay the groundwork for future research. The linguistic markers and R code used for the NLP analysis can be further developed and expanded. In particular, linguistic expressions of identity fusion should be studied further. This thesis took a conservative, evidence-based approach to generate the lists of linguistic markers for the different narrative categories, including fusion. By using exclusively fusion markers that have been identified in previous research projects (i.e. kinship language and metaphors of shared blood), the analysis might however have missed other possible linguistic expressions of fusion. Future research should also perform statistical analysis for violence risk markers observed across different datasets. While the analysis of this thesis provided a comparative assessment based on the newly developed Violence Risk Index, which is based on the statistical analysis of terrorist manifestos, the power and quality of inference of this study were impacted by the absence of a statistical comparison of the results for the different groups. Follow-up

projects could examine a larger number of groups and statistically compare the resulting scores. It would also be useful to work closely with CT professionals when refining the fusion-based violence prediction model in the future.

Practical Implications for Prevention

How can existing early warning systems be improved and complemented? What could new prevention and intervention approaches look like? The findings presented in this thesis might be valuable for the work of policymakers, security agencies, and the tech industry. This project has demonstrated how socio-psychological clues found in language can be used to assess the proneness to violence among users in ideologically extreme online groups. The new Violence Risk Index developed in this doctoral thesis can contribute to improving and expanding existing early warning systems used by law enforcement and security efforts.

Traditionally, security and intelligence services have assessed the risk of groups or individuals committing a terrorist attack based on a range of factors, including the possession of weapons, membership of listed terrorist groups, and explicit calls to violence or credible threats against concrete targets (Council of Europe, 2022; FBI, 2022; Fein & Vossekuil, 2000; HM Government, 2015). However, new technological, organisational and cultural trends in terrorist networks have rendered these indicators increasingly unhelpful.

The use of improvised self-made weapons, including 3D-printed weapons, means that the tracking of weapons has become more difficult (Veilleux-Lepage, 2021). The emergence of post-organisational online structures of violent extremist networks

makes it harder to detect individuals prone to terrorism than when they were part of centralised, locally rooted, and hierarchically organised terror groups (Davey et al., 2021). Finally, the use of satire and gamified communication by digital extremist subcultures has obfuscated traditional risk assessment approaches, which are not fit to distinguish between empty threats, trolling activities, or LARP and credible threats (Ebner, 2020; Ovide, 2021; RAN, 2021).

This thesis has presented an alternative risk assessment approach that might help tackle some of these arising challenges in the field of CT. The findings presented in this thesis suggest that it is possible to use linguistic markers to detect highly fused actors at risk of carrying out acts of violent self-sacrifice, without them being aware of it. The research therefore holds out the prospect of new forms of the early detection and prevention of terrorist violence, rooted in a robust body of theory linking fusion to violence. An important advantage of the new framework is that the examined key variables are not influenced by ideological or cultural factors. The framework can therefore be applied in different geographic and demographic contexts. As the relevant variables are revealed unconsciously in language, they also reach beyond strategically chosen words of escalation or de-escalation and are a more reliable predictor than explicit threats to violence. A way forward could be to integrate the socio-psychological markers of the Violence Risk Index, including identity fusion, into existing violence risk assessment frameworks.

As noted in the limitations above, the Violence Risk Index can, however, not replace manual investigations. False positives and false negatives in the pre-crime arena can come at an enormous cost: false positives may lead to unfair arrests or other human

rights restrictions of suspected individuals, while false negatives could entail a police failure to stop a deadly attack. Although the measure developed here can be used in an initial risk assessment to narrow down the scope of investigations and allocate human resources to high-risk groups and individuals in online spaces, manual reviews are still essential to confirm the predicted risk level.

Even individuals who show all the identified psychological signs of radicalisation towards terrorism may never resort to violence. Additional barriers can, for example, arise from high emotional or economic stakes (e.g. due to care duties for a young child or substantial financial losses for the entire family) that might cause prone individuals to reconsider violent action. The framework is also not designed to capture every single case of radicalisation towards violent extremism and terrorism. For example, the analysis of Elliot Rodger's manifesto showed that the terrorist's writings did not carry any of the expected psycholinguistic markers. Instead of fusing with an in-group and demonising a clearly defined out-group, he viewed himself as the sole victim in a constant fight against the rest of humanity. Violent misogyny, in particular cases emerging from the Incel community, may be an important exception to violence risk patterns identified in this project. One notable difference is that terrorists emerging from the misogynist Incel subculture are often experiencing mental health issues related to loneliness. The underlying drivers of some of the perpetrators of extreme violence may therefore not be found in extreme forms of group cohesion but rather in psychological and/or social pathologies.

Practical Implications for Deradicalisation

The importance of identity fusion in radicalisation pathways towards violence suggests that identity-centred approaches should be prioritised in deradicalisation and intervention programmes. Based on the findings of this thesis, policymakers and intervention providers should consider developing measures that tackle identity fusion or other statistically significant factors in the Violence Risk Index. For example, one route may consist of de-fusing highly fused individuals from their in-group. An alternative method could entail diverting at-risk individuals away from their fused (online) identities to other identities that may not be fused.

The results of this doctoral project point to the potential of fusion-based approaches to early prevention and deradicalisation. Previous researchers have proposed the creation of a new set of early prevention and intervention techniques based on the idea of de-fusion, i.e. the reversal of identity fusion (Whitehouse, 2021). Based on the known pathways to fusion, a promising route could be to focus on deconstructing perceptions of shared transformative experiences (Whitehouse et al., 2017). For example, highlighting unshared experiences among members of a group could have a disruption effect on in-group fusion dynamics, potentially allowing members to detach their own experiences and identities from the group.

As fusion is only thought to constitute a violence risk factor when combined with existential threat, out-group entitativity, and violence-condoning norms, it would also be possible to focus on reducing the levels of these other variables. Fusion might even be redirected towards a more positive pro-group cause when the other ingredients that

react toxically with fusion are removed through targeted interventions. Previous CT and countering violent extremism (CVE) initiatives have focused on reducing out-group demonisation and dehumanisation by re-humanising the perceived enemy (Shaw et al., 2022). For example, this can be done by exposing radicalised individuals to the perceived enemy group (Lankford & Gillespie, 2011) or by presenting them with tangible accounts of victims who became targets of terrorism (Neumann, 2013). An individual's adherence to violence-condoning group norms might be weakened by demonstrating peaceful alternative paths for action by casting doubt on the glorification of martyrdom or violent role models or by refuting justifications of violence.

Another way of translating the findings into practical approaches could be to focus on diversion instead of de-fusion. For example, it may be possible to target interventions at other layers of personal identities that are not fused with an extremist in-group. Members of violent and ideologically extreme groups tend to continue to maintain alternative (potentially non-fused) identities, which could be leveraged in targeted intervention. The psychiatrist Robert Lifton interviewed former Nazi doctors working in concentration camps and terrorists of the Japanese Aum Shinrikyo cult and found that it was common among these individuals to commit violence in one compartment of their lives, while maintaining a normal social life with their families and friends in a separate compartment (Lifton, 1988). This effect of "split identities", which may have been strengthened by social media and the possibility to hold multiple online and offline identities, could be used in deradicalisation efforts that make users turn away from their fused identity.

Technology has made it easier to split and multiply our identities. Split identities constitute a potential limitation for fusion-based research because they may make it difficult to capture all the identities of an individual, analysing just their language and behaviour in one virtual community. However, the idea of split identities might also provide an innovative route for deradicalisation and intervention. For example, it may be possible to divert people away from one identity to another one that may not be fused. Future research could ask questions such as: How can we spot when people around us start splitting their identities? And what can be done to make them focus on one rather than another identity?

From the outside, Alek Minassian seemed like a normal boy from Ontario. He had no criminal record and became a software and mobile app developer after enrolling in the Canadian Armed Forces. His former classmates described him as “harmless” and a senior military official said “there were no red flags”. However, online Minassian became obsessed with the misogynist Incel subculture and built a new heroised self that would protect him from feelings of weakness and worthlessness. When Alek Minassian drove his van into a crowd of pedestrians in Toronto in 2018, killing 10 people and injuring 16, he did so as a self-proclaimed Incel: “The Incels Rebellion has already begun!”, he wrote. The 25-year-old was a virgin and hated women for not sleeping with him. “I’m a supreme gentleman”, he said in an interview with the police right after he was arrested, further stating “I was angry that they would give their love and affection to obnoxious brutes”. At his trial in November 2020, he said he wanted to make it onto a website that ranks mass murderers by how many people they killed (Casey, 2020).

It is possible to maintain one or more online identities that strongly differ from one's offline identity. One can be a member of a dangerous online extremist group and continue to work in a well-respected job without anyone knowing. Identity splitting is not a result of a psychotic disorder. It can even be adaptive and healthy to separate one's private and professional identity, for example, by pursuing unusual hobbies that help compensate tedious jobs. There are entire societies in East Asia where behaving the same in different contexts would be seen as pathological (Yuki & Schug, 2012). However, split identities can also facilitate the growth of one extreme identity and enable otherwise ordinary people to do terrible things. The idea of identity splitting is as old as folklore, and (from the legend of Faust, to Jekyll and Hyde, to Lord Voldemort's Horcruxes) it is seldom seen in a good light. Nonetheless, split identities may provide new opportunities for deradicalisation and targeted intervention in the fields of CT and CVE that can be explored in future research that could draw on existing evidence from both psychological phenomena, i.e. "identity splitting" and "identity fusion".

Practical Implications for Online Interventions

This doctoral project is about the early warning signs of extreme violence and their linguistic identification in digital spaces rather than harmful online content in the wider sense. However, the findings of the conducted research have implications for online interventions and merit a discussion of the different approaches in response to extremist content on online platforms. Specifically, this thesis' cases studies showed that extremist online groups can create and foster an atmosphere in which violence risk indicators such as violence condoning norms and the dehumanisation of out-groups are promoted. Several options exist to address extremist activities, users and content in

online spaces: regulation and removal laws, content moderation and reporting mechanisms, content deprioritisation and algorithmic changes, education and counter-communication.

In the past few years, new legal frameworks were passed to regulate online hatespeech and harmful content. The German government's anti-hatespeech law "Netzwerkdurchsuchungsgesetz" (Network Enforcement Act) of 2017 (BMJ, 2017) and the EU's Digital Services Act of 2022 (European Commission, 2022) are examples of nationally or transnationally imposed laws that have entered into force in Europe. At the time of writing, the UK's Online Safety Bill 2023 (UK Parliament, 2023) is being reviewed by the House of Lords. These laws include clauses that require the removal of accounts or pieces of content that are inciting hatred, defaming individuals or glorifying terrorism. However, a significant fraction of potentially harmful content is not illegal. Many extremists operate in the legal grey zones to avoid take-downs of their content. For example, hateful content camouflaged as satire or artistic self-expression can feed into an atmosphere of dehumanisation and violence endorsement in which violent extremism can thrive (Fielitz & Ahmed, 2021; Ebner, 2020).

There is an ongoing debate in academia as well as in policy circles about whether removal policies that target harmful content and accounts that promote radicalisation are the most efficient or even the most effective solution (Pearson, 2018; Chandrasekharan et al, 2017; Clifford, 2021; Wallner, 2021). Researchers have compared the removal of content and the deplatforming of users with a cat-and-mouse game, whereby slightly altered content and accounts continue to replace the ones that were taken down (Bodó, 2018, Clifford & Powell, 2019). This dynamic casts doubt on the

efficiency of removal and suspension policies. However, other studies suggested that suspensions of extremist accounts have curbed the number of replies, retweets and overall dissemination (Berger & Morgan, 2016; Berger & Perez, 2016; Alexander, 2017). An in-depth study of the consequences of disruptions and removals of ISIS channels on Telegram found that these online interventions had a profound and long-term impact on the number of jihadist posts and active jihadist accounts on the messaging app. The researchers concluded that take-downs can be effective when performed in a coordinated, sustained manner and combined with offline interventions (Amarasingam et al, 2021). Berger and Morgan (2016, p. 15) have compared the necessity of take-downs to weeding: "The consequences of neglecting to weed a garden are obvious, even though weeds will always return."

Yet, there are also potentially negative side effects of removal policies on social media platforms. For example, account suspensions and content removals tend to prompt extremist groups and networks to move to less visible, more isolated online spaces, which are currently not included in laws such as the Network Enforcement Act (Wallner, 2021; Guhl et al, 2020). The emergence of alt-tech platforms has been documented as a direct consequence of online removal policies introduced on big technology platforms by some researchers. The associated online migration towards more hidden spaces can make it harder for law enforcement agencies, security professionals and researchers to detect, monitor and analyse extremist activities (Guhl et al, 2020; Ebner, 2020).

Another adverse effect is that removal policies can spark and nourish grievances that might further fuel anti-establishment conspiracy myths and extremist ideologies.

Pearson warned that repressive measures such as take-downs might reinforce extremist online communities due to shared negative experiences and emotions (Pearson, 2018). Other researchers have raised concerns about perceived double standards and inconsistent policies, which can strengthen far-right extremist victimisation narratives about unjustified censorship and free speech infringements (Allchorn, 2020; Klein, 2020). For example, grievances about removal policies have been exploited by extremist groups who portray governments and big technology firms as complicit in establishing an Orwellian surveillance state (Ebner, 2023).

Beyond legally imposed removal requirements, other potential online intervention routes are content moderation, deprioritisation, digital education and counterspeech initiatives. Content moderation may involve the monitoring and assessment of user-generated content by dedicated teams of an online platform provider to identify extremist posts and take appropriate action. The big technology firms tend to use a combination of AI models and manual reviews to detect posts that violate their platform policies. Platform moderation can be combined with reporting mechanisms, allowing users to flag or report content that violates a platform's community standards. Most large platforms such as Facebook, Instagram and Youtube have reporting mechanism in place that enable them to take action against flagged accounts or content such as removal, suspension or blocking (Meta, 2023; YouTube, 2023).

Platforms may consider deprioritising content or helping to counter it rather than removing specific posts. Fostering digital literacy and awareness about extremist manipulation strategies used online can empower users to respond themselves. For

example, tools and resources might teach them the skills needed to better protect themselves and others from targeted hatred and harassment. Such skills may range from knowing how to block or report specific accounts and how to change privacy settings to recognising signs of radicalisation in others, seeking psychological support or launching an effective counter-communication campaign. A growing number of think tanks and civil society organisations are already involved in building resilience to online extremism and hatespeech. The German organisation HateAid offers legal and psychological help for individuals targeted by or at risk of online harassment and hate (Hate Aid, 2023), while the ISD has designed a digital citizenship curriculum for formal and informal education settings (ISD, 2023).

Ultimately, it is necessary to look at the algorithmic architecture and underlying business models of technology firms such as Google, Meta and Tik Tok. Social media platforms tend to be designed to harvest human attention and therefore inadvertently prioritise content that is extreme or sensationalist (Aral, 2020). For example, studies showed that video sharing platforms such as YouTube behave like rabbit holes that take users to increasingly extreme content due to the recommendation algorithm that seeks to maximise click rates and user engagement (Tufekci, 2018; Brown et al.,2022). This implies that platform regulations will only be sustainable, if they also include legal requirements for companies to change their algorithms to avoid amplifying radicalisation processes.

Ethical Limits and Lessons from the Past

Putting the research findings of this thesis into praxis bears the dangers of undermining user privacy and criminalising socio-psychological phenomena. Many ethical limits have been crossed in the name of the prevention of terrorism (Silke, 2019). Looming ethical challenges related to new detection and prevention systems need to be addressed and informed by lessons learnt from past CT and CVE approaches.

It is important to address pressing questions related to the protection of user privacy. How far can early warning systems go without posing a threat to our fundamental freedoms and our right to privacy? What are the limits for digital analytics and preventive policing and how can bias be avoided? Careful navigation through these questions of how to balance freedom and security will be needed to develop sustainable, effective, and ethical future prevention and intervention mechanisms.

Policymakers and law enforcement agencies need to carefully weigh off data privacy concerns with gains for national security. While big data collections are needed to train AI algorithms and can be a powerful tool for crime prevention, they also bear the risk of being exploited for political purposes and infringing on privacy rights. As this research project has demonstrated, it may be possible to protect user anonymity by carrying out initial risk assessments based on anonymised data. Different countries have different legislative frameworks regarding data security and the possible applications of new technologies. While the EU takes a more protective and risk-averse approach towards digital technologies, as exemplified by the GDPR and the EU AI Act, the US and

the UK have tended to adopt more liberal views on the use of big data and technological innovations (Beens, 2020; Roberts & Floridi, 2021).

Another key risk related to pre-crime monitoring systems stems from the potential profiling of specific communities, which could be disproportionally targeted in future violence risk assessments. The new framework developed and presented in this thesis is ideologically and ethnically agnostic. Nonetheless, it entails the gathering of psychological knowledge about online users, and that knowledge could be misused. It is therefore important to consider the limits for the applications of this research and its reconciliation with the protection of privacy and other fundamental rights.

Many measures introduced in the aftermath of 9/11 and the 7/7 London tube bombings to prevent and counter future terrorist activities have led to a systematic profiling of ethnical, religious, or cultural minority groups, which has in turn has fuelled community tensions, minority group grievances, collective identity crises, and other societal factors that may in the long-run prove counter-productive to reducing the terrorism threat (Metropolitan Police Service, 2008; Ministry of Justice, 2009). In the UK context, it is worth looking at the Stop and Search Programme as an early warning system that made use of profiling. However, other related pre-crime measures include Pre-Charge Detention, the Control Orders, and the Terrorism Prevention and Investigation Measures (TPIMs). Various studies have shown that these policies have disenfranchised and alienated minority communities that felt unjustly targeted by CT policies (Lister & Jarvis, 2013; Parmar, 2011).

Data collected by the Metropolitan Police Service from across London Boroughs for the 2008 Equality Impact Assessment indicated a racially disproportionate use of Stop and Search powers (Metropolitan Police Service, 2008). Statistics published by the Ministry of Justice back in 2013 supported the claim that Stop and Search practices had a disproportionate impact on Asian and Black ethnic minorities. Under Section 1 PACE powers, Black persons were six times and Asian persons were over two times more likely to be searched than white persons (Ministry of Justice, 2009). Human rights and civil liberty rights groups such as Liberty, Amnesty, and the Joint Committee on Human Rights are not the only ones that have warned of the negative effects of such profiling-based approaches on community cohesion and societal resilience to extremism. Reports by the former independent reviewer of terrorism legislation, Lord Carlile of Berriew, and evidence collected for the Home Office's Full Equality Impact Assessment reached similar conclusions about the dangers of vaguely restricted policing powers (Home Office, 2011).

Both qualitative and survey research have found that there is a clear negative correlation between an individual's experience of being stopped and searched and their confidence in the police. In particular, the disproportionate targeting of ethnic minorities has contributed to hostile attitudes and less respect towards the police in these communities (Miller et al., 2000). This claim is supported by academic assessments of the enforcement of profiling-related police powers, which have also been found to have a disproportionate collateral impact on ethnic minorities. For example, Ben Bowling and Coretta Phillips's research showed that this "racial effect" has harmed

community cohesion, damaged police–community relations, and undermined the police's legitimacy (Bowling & Philips, 2007).

Alpa Parmar's research suggested that the high disproportionality rates are "inherently problematic" and counterproductive to national security aims (Parmar, 2011). Section 44 has led to the perceived criminalisation of ethnic minority communities, the compromising of civil liberties, and the worsening of inter-ethnic tensions. Parmar's empirical study found that Stop and Search has created grievances and triggered resentment among ethnic and religious minorities. For example, many Asians and Blacks interviewed by Parmar indicated that they felt subject to speculative intrusion and ethnic profiling. One British-Pakistani Muslim told Parmar: "That's the third time I've been stopped and searched. What, do I have terrorist written on my forehead? They think it's OK to brand us as criminals and that we'll just sit back (…)." Another Bangladeshi interviewee indicated that his Stop and Search experience made him more suspicious of the police and less confident about the fairness of society (Parmar, 2011).

Lister and Jarvis (2013) focused their study on the impact of Stop and Search on questions related to identity and citizenship. Their focus groups explicitly voiced opposition to the existence and implementation of Stop and Search powers, accusing the policy of targeting "Asian faces" and "creating racism". The qualitative study concluded that the policy has had a negative impact on ethnic minority communities' experience of citizenship. In particular, Black and Asian citizens have "experienced an erosion of rights, a dampening of political engagement and a weakening sense of attachment to British identity" (Lister & Jarvis, 2013, p. 759).

Finally, it is important to consider potential areas for the exploitation of violence risk assessment frameworks, such as the one presented in this thesis. Human rights infringements and the strategic criminalisation of specific population segments in the name of CT has been a key concern for the UN (United Nations Office on Drugs and Crime [UNODC], 2023). The absence of an international definition of "terrorism" may have facilitated the misuse of the term in autocratically governed countries to prosecute political dissidents and minority communities who have not been involved in terrorist or even criminal activities (UNODC, 2023). When developing new computationally supported risk assessment methods, including based on this project's findings, intelligence and CT professionals should therefore carefully consider the integration of safeguarding mechanisms to avoid terrorism prevention approaches developed in democratic systems being easily repurposed for non-democratic aims.

To conclude, the practical implications of this thesis's findings should be treated with caution. Even though the risk assessment framework presented in this thesis looks at psychological patterns that can be found across the entire spectrum of society, it could be exploited to target specific ethnic, religious, or cultural communities. As with any pre-crime measures, a careful approach is needed to avoid reinforcing old forms of ethnic, religious, or cultural profiling and prevent the rise of new forms of psychological profiling. The same methodology should be applied to groups across the entire political spectrum to understand whether being fused to a revolutionary ideology is a risk factor regardless of the nature of the ideology. One promising application of the research might consist in narrowing down the scope of security investigations to distinguish between empty words and credible threats. However, the Violence Risk Index should

not be used to produce a final judgement about an individual's or group's violence intentions. Instead, a high violence risk score needs to be followed up with rigorous manual investigations. Criminalising socio-psychological processes would be just as dangerous as policing ideologies. It is not only our fundamental right to think what we want, we should also be allowed to feel what we want.

The Future of Research, Policy, and Practice

AI and the metaverse are set to bring the next big wave of disruptions to CT. As early adopters of technology, violent extremists are likely to instrumentalise these innovations to fit their own purposes. How does AI change the future of terrorism and CT? What happens when virtual communities that resemble reality more than ever before are exposed to violent or hateful content? What does the emergence of these new technologies mean for early detection, prevention, and policing? How far can security services go in their use of AI in prevention and intervention? And what should the role of law enforcement agencies be in the metaverse?

Extremists have been historically good at exploiting technological innovations and using them to enhance their recruitment, propaganda, and attack tactics (Ebner, 2019). The malicious use of AI technologies is a rapidly evolving threat to the safety, democratic integrity, and political stability of liberal democratic countries. This has created unprecedented national security challenges faced by the policy community, requiring policymakers, industry representatives, and researchers to work together to update existing defence systems and to create new forecast, prevention, and response mechanisms (Brundage et al., 2018). The democratisation of AI technologies has

lowered the barriers for non-state actors, such as jihadists, extreme-right accelerationists, or anti-democracy extremists, to stage high-impact attacks in an effort to provoke radical political or societal change (Kirschke-Schwartz & Clarke, 2021; Kreps, 2021). For example, AI-supported tools can be used to launch physical attacks, cyber-attacks and information operations. Threats related to the malign use of AI technologies by non-state actors may arise both from outside the target country, sometimes backed by foreign state actors, and from within the target country in the form of homegrown extremism (Brundage et al., 2018; Kreps, 2021).

In the future, the potentially rising sophistication of terrorist activities will need to be mirrored in more technically advanced forecast and prevention mechanisms. Today, AI-based predictive policing is still primarily used for geographic hot-spot mapping and spatial risk calculations. However, risk assessments of individual offenders, in particular in relation to re-offense risk, have seen strong growth in the policing field in recent years (Babuta, 2017; Babuta & Oswald, 2019). Research by the RUSI showed that human bias can be replicated and even amplified by AI-supported predictive policing (Babuta & Oswald, 2019). As mentioned above, Black people and other ethnic minorities are disproportionally targeted by police action. If the algorithms are based on police data, they will reproduce patterns of human profiling as reflected in the datasets used to train them. This means that human biases will be passed on to AI-supported systems that will create biased risk predictions to inform future police forces action, hence creating a feedback loop that becomes a self-fulfilling prophecy (Babuta & Oswald, 2019).

Innovations in the field of AI can enable terrorists to increase their attack frequency, lethality, and terror instilled in the population. For example, AI technologies may be used to detect defence vulnerabilities in high-stake targets, as well as to maximise the damage caused in a single attack. This might make it more likely for attacks on minority communities or democratic institutions to attain the desired effects of exacerbating fears, widening societal divisions, and intimidating political adversaries. Commercially available AI technologies can be used both in the planning and execution stages of terrorist attacks. For instance, an unmanned smart drone could be employed to gather information about the target terrain (as the Christchurch shooter demonstrated), to drop explosive devices (or, in an unlikelier scenario, chemical, biological, radiological, and nuclear [CBRN] weapons), or to livestream the attack.

Beyond real-world attacks, threats also emerge from the potential AI-supported identification of vulnerabilities in critical infrastructure and automated hacking, as well as from the security vulnerabilities of AI-based systems that are used in public and private sector organisations. If AI-based systems are embedded within critical infrastructure, this could provide high-impact targets for black hat hackers, whether state or non-state backed. AI-powered technologies, such as self-driving cars or unmanned public transport, tend to be prone to hacking attacks, which could in turn pose a significant risk to national security.

New AI-powered technologies may also be used in information operations. For example, AI-based text generation and image manipulation software can be harnessed to produce and spread pieces of disinformation (Brundage et al., 2018; Kreps, 2021). By using AI-powered speech synthesis or impersonation, it is possible to create credible

content that can then be disseminated by automated or semi-automated trolling armies (Brundage et al., 2018; Kreps, 2021). AI-based language tools that can be applied to virtual agents and automated chats may be a more concerning trend. These technologies are very close to passing the Turing test and they are commercially available at low cost (Kreps, 2021). Even before the latest generation of machines that sound like humans, ISIS recruiters used bots and automated messages that targeted users with Muslim sounding names and British phone numbers on Telegram (Ebner, 2020). In the future, AI technologies may enable terrorists, extremists, and trolling armies to shift the power asymmetry in the online and offline battlefields in their favour.

There is also enormous potential for AI to be used in the security sector to predict and counter the activities of violent extremists and terrorists. As argued in this thesis, early detection mechanisms based on violence predictors found in language are becoming more sophisticated, which can lead to significant improvements in the field of terrorism prevention. AI-based technologies, combined with big data processing capabilities, have also been driving the development of smarter, more targeted, and more reactive early warning and defence systems. Significant improvements have been made in areas such as threat monitoring, combat simulation, and target recognition (AI World School, 2023). Responding to AI-related national security challenges on the horizon will require the policy community to invest in R&D as well as legal innovation and public awareness campaigns. It will be important to integrate safe-by-design mechanisms into new commercially available tech products that protect them from potential misuse for radical political purposes, such as the anti-forgery function of commercial photocopiers that prevents their users from copying money (Gray, 2022).

The policy community will need to work closely with researchers, as well as the developers of cutting-edge AI technologies, to forecast, prevent, and mitigate their potential exploitation.

Likewise, the metaverse and virtual reality (VR) innovations can be weaponised by extremists. For instance, terrorist attacks may be livestreamed to audiences in a 3D VR setting. In the metaverse, tens of thousands of people can be in the same virtual environment. VR-based terrorism can increase the amount of fear generated by terrorist attacks, as well as blurring the lines between reality and fiction and inspiring copycat attacks. New CT challenges arising from the metaverse and VR-supported technologies will need to be addressed in future research, policy, and practice. Policymakers and law enforcement agencies will face questions such as: Should intelligence and police units establish a presence in the metaverse? How can they best assess content generated in new VR environments? And how far should they go in interacting with users in the metaverse who are showing signs of radicalisation towards violence?

As with AI-based technologies, VR also carries an enormous potential for future terrorism prevention and intervention approaches. For example, it might be used for immersive awareness-raising campaigns. Individuals at risk of radicalisation may participate in VR scenarios that simulate scenarios of discrimination or hate from the perspective of the targeted victim. Such VR-based tools are already used in policing training programmes to tackle bias and discrimination in the police force (Alanis & Pyram, 2022). Another potential area for VR application in radicalisation prevention is the use of holograms and VR recordings of eye-witnesses who document their first-hand

accounts of historic events, such as World War II, to prevent phenomena such as Holocaust denial among future generations.

To conclude, VR and AI technologies are creating a new set of interlinked challenges for national security. They have expanded existing threats and given rise to new threats both in the digital and physical realm. When powered by AI or amplified by VR, both cyberattacks and physical plots can be more scalable, more targeted, and more difficult to attribute. Beyond short-term national security threats, they can also exacerbate the ongoing disinformation crisis, widen societal divisions, and decrease trust in democratic, scientific, and media institutions. These threats on the softer end of the national security spectrum may arguably pose a more significant long-term challenge than terrorist attacks. However, they can also provide new windows of opportunity for prevention and intervention.

Bibliography

Abbas, T. (2022). *The Buffalo attack – An analysis of the manifesto.* The Hague: International Centre for Counter-Terrorism.

Ackerman, G., & Peterson, H. (2020). Terrorism and COVID-19: Actual and potential impacts. *Perspectives on Terrorism, 14*(3), 59–73.

Agre, P. (1999). Life after cyberspace. *European Association for the Study of Science and Technology (EASST), 18*, 3–5.

Ahearn, E.-R., Bhui, K., & Jones, E. (2020). What factors are truly associated with risk of radicalisation? A secondary data analysis within a UK sample. *Transcultural Psychiatry, 58*(5) 645–653.

AI World School. (2023, 11 March). *Scope of artificial intelligence in military operations*. https://aiworldschool.com/research/scope-of-artificial-intelligence-in-military-operations/ (retrieved 11 March 2023).

al-Qaradawi, Y. (1960). *The lawful and prohibited in Islam.* Davao: Wisdom International School for Higher Education Studies (WISHES).

Alanis, J. M., & Pyram, R. H. (2022). From simulations to real-world operations: Virtual reality training for reducing racialized police violence. *Industrial and Organizational Psychology, 15*(4), 621–625.

Alexander, J. (2018, 28 February). *Discord is purging alt-right, white nationalist and hateful servers.* Polyglon.

Alexander, A. (2017). *Digital Decay: Tracing change over time among English-language Islamic State sympathizers on Twitter.* Washington D.C.: The George Washington University, Program on Extremism.

Allan, H., Glazzard, A., Jesperson, S., Reddy-Tumu, S., & Winterbotham, E. (2015). *Drivers of violent extremism: Hypotheses and literature review*. London: Royal United Services Institute (RUSI).

Allchorn, W. (2020). *Free speech defenses and the far right: The cases of Tommy Robinson, Geert Wilders and Milo Yiannopoulos.* Europe Now Journal. 15 January 2020.

Allen, C. (2019). *National action: Links between the far right, extremism and terrorism*. Leicester: UK Commission for Countering Extremism.

Alon, N., & Omer, H. (2005). *The psychology of demonization: Promoting acceptance and reducing conflict.* London: Routledge.

Amanullah, Z., & Harrasy, A. (2017). *Between two extremes: Responding to Islamist and tribalist messaging online in Kenya during the 2017 elections*. London: Institute for Strategic Dialogue.

Amarasingam, A., & Argentino, M.-A. (2020). The QAnon conspiracy theory: A security threat in the making. *CTC Sentinel, 13*(7), 37–44.

Amarasingam, A., Maher, S. & Winter, C. (2021). *How Telegram Disruption Impacts Jihadist Platform Migration.* London: Centre for Research and Evidence on Security Threats.

Anderson, B. (1983). *Imagined communities: Reflections on the origin and spread of nationalism.* London: Verso.

Anderson, L., & Lepore, E. (2013). Slurring words. *Nous, 47*, 25–48.

Applebaum, A., Pomerantsev, P., & M. Smith, C. C. (2017). *“MAKE GERMANY GREAT AGAIN”: Kremlin, alt-right and international influences in the 2017 German elections*. London: Institute for Strategic Dialogue and London School of Economics (LSE) Arena.

Aral, S. (2020). *The Hype Machine: How social media disrupts our elections, our economy and our health – and how we must adapt.* New York: Harper Collins.

Argentino, M.-A., & Crawford, B. (2021). *The WQmen of QAnon*. London: Global Network on Extremism and Technology.

Arifka, D., Hakim, M. N., Adhipta, A. S., Yogananda, K. S., Salsabila, R., & Ferdiana, R. (2022). Pandemic fatigue: An analysis of Twitter users’ sentiments against the COVID-19 in Indonesia. *Jurnal Psikologi Volume, 49*(2), 182–194.

Aron, A., E. Aron, E., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*, 596–612.

Ashmore, R., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin, 130*(1), 80–114.

Association of Internet Researchers [AOIR]. (2020). *Ethics*. https://aoir.org/ethics/ (retrieved 15 June 2021).

Assoudeh, E., & Weinberg, L. (2021). *In Western Europe, right wing terrorism is on the rise*. Open Democracy.

Atran, S. (2016). The devoted actor: Unconditional commitment and intractable conflict across cultures. *Current Anthropology, 57*, 192–201.

Avaaz, & Institute for Strategic Dialogue [ISD]. (2019). *Disrupted: Evidence of widespread digital disruption of the 2019 European elections*. London: Avaaz and Institute for Strategic Dialogue.

Awan, A., Hoskins, A., & O’Loughlin, B. (2011). *Radicalisation and media: Connectivity and terrorism in the new media ecology.* London: Routledge.

Ayad, M., Amarasingam, A., & Alexander, A. (2021). *The cloud caliphate: Archiving the Islamic State in real time*. West Point, NY: Institute for Strategic Dialogue and Combating Terrorism Center at West Point.

Babuta, A. (2017). *Big data and policing: An assessment of law enforcement requirements, expectations and priorities*. London: Royal United Services Institute for Defence and Security Studies.

Babuta, A., & Oswald, M. (2019). *Data analytics and algorithmic bias in policing*. London: Royal United Services Institute for Defence and Security Studies.

Bäck, E., Bäck, H., Sendén, M., & Sikström, S. (2018). From I to we: Group formation and linguistic adaption in an online xenophobic forum. *Journal of Social and Political Psychology, 6*(1), 76–91.

Bailey, G., & Edwards, P. (2017). Rethinking ‘radicalisation’: Microradicalisations and reciprocal radicalisation as an intertwined process. *Journal for Deradicalization, 10*, 255–281.

Baillet, S. (2019). *A short pre-action report*. Terrorist manifesto, Halle.

Baker, R. W. (2003). *Islam without fear: Egypt and the new Islamists.* Cambridge, MA: Harvard University Press.

Baker, S. (2022). Alt. health influencers: How wellness culture and web culture have been weaponised to promote conspiracy theories and far-right extremism

during the COVID-19 pandemic. *European Journal of Cultural Studies, 25*(1), 3–24.

Bale, J. (2007). Political paranoia v. political realism: On distinguishing between bogus conspiracy theories and genuine conspiratorial politics. *Patterns of Prejudice, 41*, 45–60.

Ball, P. (2012). *The transformation of war: Modelling modern conflict.* Berlin: Springer.

Bandura, A. (2004). The origins and consequences of moral disengagement: A social learning perspective. In F. Moghaddam, & A. J. Marsella (Eds.), *Understanding terrorism: Psychosocial roots, consequences, and interventions*. Washington DC: American Psychological Association.

Bansal, S., Vishal, G., Suhane, A., Patro, J., & Mukherjee, A. (2020). *Code-switching patterns can be an effective route to improve performance of downstream NLP applications: A case study of humour, sarcasm and hate speech detection*. Toronto: Association for Computational Linguistics.

Barnett, B. A. (2015). 20 years later: A look back at the Unabomber manifesto. *Perspectives on Terrorism, 9*(6), 60–71.

Bartlett, J. (2018). *The People vs Tech: How the internet is killing democracy (and how we save it).* London: Ebury Press.

Becker, D., Troschke, H., & Allington, D. (2021). *Decoding antisemitism: An AI-driven study on hate speech & imagery online*. Berlin: TU Berlin.

Beckett, L. (2020, 16 October). QAnon: A timeline of violence linked to the conspiracy theory. *The Guardian*.

Beelmann, A. (2020). A social-developmental model of radicalization: A systematic integration of existing theories and empirical research. *International Journal of Conflict and Violence, 14*(1), 1–14.

Beens, R. E. (2020, 29 July). The privacy mindset of the EU vs. the US. *Forbes*.

Berger, J.M., & Morgan, J. (2015). *The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter.* The Brookings Project on U.S. Relations with the Islamic World. Analysis Paper 20.

Berger, J.M., & Perez, H. (2016). T*he Islamic State's diminishing returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters.* Washington D.C.: The George Washington University, Program on Extremism, Occasional Paper.

Berger, J. (2017). *Extremist construction of identity: How escalating demands for legitimacy shape and define in-group and out-group dynamics*. The Hague: International Centre for Counter-Terrorism.

Berger, J. (2018). *Extremism*. Cambridge, MA: MIT Press.

Beutel, A. (2019). *The New Zealand terrorist's manifesto: A look at some of the key narratives, beliefs and tropes*. College Park, MD: START.

Bilik, M., Hekimoglu, E., & Gencoz, F. (2021). Traces of the unconscious in language. *Language and Psychoanalysis, 10*(1), 1–9.

Bjørgo, T., & Ravndal, J. A. (2019). *Extreme-right violence and terrorism: Concepts, patterns, and responses*. The Hague: International Centre for Counter-Terrorism.

Bjorgo, T., Thomassen, G., Strype, J., & Jupskas, A. R. (2022). Patterns and consequences of threats towards politicians. *Perspectives on Terrorism, 16*(6), 100–119.

Blom, A. (2011). Do jihadist 'martyrs' really want to die? An emic approach to self-sacrificial radicalization in Pakistan. *Revue Française de Science Politique, 61*, 867–891.
Blout, E., & Burkart, P. (2020). White supremacist terrorism in Charlottesville: Reconstructing 'Unite the Right'. *Studies in Conflict & Terrorism*. https://doi.org/10.1080/1057610X.2020.1862850
Bluteau, J. M. (2019). Legitimising digital anthropology through immersive cohabitation: Becoming an observing participant in a blended digital landscape. *Ethnography, 22*(2), 1–19.
Bodó, L. (2018). *Decentralised terrorism: The next big step for the so-called Islamic State.* VOX-pol. https://www.voxpol.eu/decentralised-terrorism-the-next-big-step-for-the-so-called-islamic-state-is/ (retrieved 15 June 2023).
Boellstorff, T. (2008). *Coming of age in Second Life: An anthropologist explores the virtually human.* Princeton, NJ: Princeton University Press.
Boellstorff, T. (2016). For whom the ontology turns: Theorizing the digital real. *Current Anthropology, 57*(4), 387–407.
Boellstorff, T., Nardie, B., Pearce, C., & Taylor, T. (2012). *Ethnography and virtual worlds: A handbook of method.* Princeton, NJ: Princeton University Press.
Bordieu, P., & Colemen, J. (1991). *Social theory for a changing society.* Boulder, CO: Westview.
Bowling, B., & Philips, C. (2007). Disproportionate and discriminatory: Reviewing the evidence on police Stop and Search. *The Modern Law Review, 70*(6), 936–961.
Boyd, R. L., & Schwartz, H. A. (2020). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology, 40*(1), 21–41.
Brandwatch. (2018, 4 October). *A defining moment: Crimson Hexagon is joining Brandwatch*. https://www.brandwatch.com/blog/brandwatch-and-crimson-hexagon/ (retrieved 3 December 2019).
Breese, E. B. (2015). When marketers and academics share a research platform: The story of Crimson Hexagon. *Journal of Applied Social Science, 10*(1), 3–7.
Breivik, A. B. (2011). *2083 – A European Declaration of Independence*. Terrorist Manifesto, Oslo.
Brewer, M. (2001). The many faces of social identity: Implications for political psychology. *Political Psychology, 22*(1), 115-125.
Brezar, A. (2022). From anti-vax to pro-Putin, how conspiracy theorists are backing Russia's war in Ukraine. *Euronews*.
Bright, J. (2017). *Explaining the emergence of echo chambers on social media: The role of ideology and extremism*. Oxford: Oxford Internet Institute.
British Psychological Society. (2018). *Code of ethics and conduct*. https://www.bps.org.uk/news-and-policy/bps-code-ethics-and-conduct (retrieved 2 October 2019).
Brown, E. (2017, 17 January). Crimson Hexagon stores one trillion social media posts for customer analytics. *ZDNet*.
Brown, M.A., Bisbee, J., Lai, A., Bonneau, R., Nagler, J, & Tucker J.A. (2022). *Echo chambers, rabbit holes, and algorithmic bias: How YouTube recommends content to real users.* http://dx.doi.org/10.2139/ssrn.411905 (retrieved 10 June 2023).

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B, Anderson, H, Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S. … Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Cambridge: Centre for the Study of Existential Risk.

Buhrmester, M., Fraser, W., Lanman, J., Whitehouse, H., & Swann, W. (2014). When terror hits home: Identity fused Americans who saw Boston bombing victims as 'family' provided aid. *Self & Identity, 14*(3), 253–270.

Buhrmester, M. D., Zeitlyn, D., & Whitehouse, H. (2020). Ritual, fusion, and conflict: The roots of agro-pastoral violence in rural Cameroon. *Group Processes & Intergroup Relations, 25*(1), 298–311.

Bundesamt für Verfassungsschutz. (2021, 29 April). *Neuer Phänomenbereich "Verfassungsschutzrelevante Delegitimierung des Staates".* Berlin.

Bundesministerium der Justiz [BMJ] (2017). *Regeln gegen Hass im Netz - das Netzwerkdurchsuchungsgesetz.* Berlin: BMJ. https://www.bmj.de/DE/Themen/FokusThemen/NetzDG/NetzDG_node.html (retrieved 3 June 2020).

Cai, W., & Landon, S. (2019, 3 April). Attacks by White extremists are growing. So are their connections. *The New York Times*.

Calderón, F. H., Balani, N., Taylor, J., Peignon, M., Huang, Y.-H., & Chen, Y.-S. (2021). Linguistic patterns for code word resilient hate speech identification. *Sensors, 21*(23), e7859.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9*(2), 48–57.

Cameron, D. (2011, 3 May). *Statement to House of Commons HC Deb. cc 461, 473*. London: HM Government.

Cameron, J. (2004). A three-factor model of social identity. *Self and Identity, 3*(3), 239–262.

Cartwright, D., & Zander, A. (1968). *Group dynamics: Research and theory.* New York, NY: Harper & Row.

Casey, L. (2020, 7 December). Alek Minassian would tell his victims he was lonely and angry at society, court told. *Global News*.

Center for the Analysis of Social Media [CASM]. (2016a, 26 January). *Demos and the Irish election*. Demos. https://demosuk.wpengine.com/project/ie/ (retrieved 3 December 2019).

Center for the Analysis of Social Media [CASM]. (2016b, 26 May). *New Demos study reveals scale of social media misogyny*. https://demos.co.uk/press-release/staggering-scale-of-social-media-misogyny-mapped-in-new-demos-study/ (retrieved 3 December 2019).

Central University Research Ethics Committee [CUREC]. (2019). *Internet-based research (IBR).* University of Oxford Research Support. https://researchsupport.admin.ox.ac.uk/files/bpg06internet-basedresearchpdf (retrieved 15 October 2020).

Chandrasekharan, E., Pavalanathan, U., Srinivasan, A. Glynn, A., Eisenstein, J. & Gilbert, E. (2017). You Can't Stay Here: The efficacy of Reddit's 2015 ban examined through hate speech. *Proceedings of the AMC on Human-Computer Interaction 1*(31), 1-22.

Chebrolu, E. (2020). The radical lens of Dylann Roof: Racial anxiety and white nationalist rhetoric on new media. *Review of Communication, 20*(1), 47–86.
Choi, V. K., Jackson, J. C., & Gelfand, M. J. (2018). The role of entitativity in perpetuating cycles of violence. *Behavioral and Brain Sciences, 41*, e196.
Choudhury, T. (2007). *The role of Muslim identity politics in radicalisation.* London: HM Government Report Commissioned by the Preventing Extremism Unit.
Clifford, B. (2021). *Moderating Extremism: The State of Online Terrorist Content Removal Policy in the United States.* Washington D.C.: The George Washington University, Program on Extremism.
Clifford, B. & Powell, H. (2019). *Encrypted Extremism: Inside the English-Speaking Islamic State Ecosystem on Telegram.* Washington D.C.: The George Washington University, Program on Extremism.
Coats, S., Smith, E. R., Claypool, H. M., & Banner, M. J. (2000). Overlapping mental representations of self and in-group: Reaction time evidence and its relationship with explicit measures of group identification. *Journal of Experimental Social Psychology, 36*, 304–315.
Cohen, K., Johansson, F., Kaati, L., & Mork, J. C. (2014). Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence, 26*(1), 246–256.
Coleman, E. G. (2010). Ethnographic approaches to digital media. *Annual Review of Anthropology, 39*, 487–505.
Colliver, C., Pomerantsev, P., & Anne Applebaum, J. B. (2018). *Smearing Sweden: International influence campaigns in the 2018 Swedish election.* London: Institute for Strategic Dialogue and London School of Economics and Political Science (LSE) Arena.
Coolsaet, R., & Renard, T. (2022). *What the Zeitgeist can tell us about the future of terrorism*. The Hague: International Centre for Counter-Terrorism.
Council of Europe. (2022). *Types of cyberviolence*. Cyberviolence at a glance. https://www.coe.int/en/web/cyberviolence/types-of-cyberviolence (retrieved 21 January 2023).
Crawford, J., Ebner, J., & Hasan, U. (2018). The balanced nation: Addressing the challenges of Islamist and far-right extremism in the classroom. In F. Panjwani, L. Revell, & M. D. R. Gholami (Eds.), *Education and Extremisms.* London: Routledge.
Crossley, M. L. (2000). Narrative psychology, trauma and the study of self/identity. *Theory and Psychology, 10*, 527–546.
Crusius, P. (2019). *The inconvenient truth*. Terrorist manifesto, El Paso.
Crystal, D. (2001). *Language and the internet.* Cambridge: Cambridge University Press.
Dafaure, M. (2020). The “great meme war:” The alt-right and its multifarious enemies. *Angles, 10*, 1–10.
D’Andrade, R. (1995). Moral models in anthropology. *Current Anthropology, 36*(3), 399–408.
Data and Society Research Institute. (2019). https://datasociety.net/ (retrieved 16 January 2023).
Davey, J. (2021). *Gamers who hate: An introduction to ISD’s gaming and extremism series*. London: Institute for Strategic Dialogue.

Davey, J., Comerford, M., Guhl, J., Baldet, W., & Colliver, C. (2021). *A taxonomy for the classification of post-organisational violent extremist & terrorist content*. London: Institute for Strategic Dialogue.

Davey, J., & Ebner, J. (2017). *The fringe insurgency: Connectivity, convergence and mainstreaming of the extreme right*. London: Institute for Strategic Dialogue.

Davey, J., & Ebner, J. (2019). *The great replacement: The violent consequences of mainstreamed extremism.* London: Institute for Strategic Dialogue.

de Beauvoir, S. (2011). *The second sex.* (English ed.). New York, NY: Penguin Random House.

Ding, F., Ge, Q., Jiang, D., Fu, J., & Hao, M. (2017). Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach. *PLoS ONE, 12*(6), e0179057.

Diplo. (2023, 12 March). *Different prefixes, same meaning: cyber, digital, net, online, virtual, e-*. https://www.diplomacy.edu/blog/different-prefixes-same-meaning-cyber-digital-net-online-virtual-e/ (retrieved 12 March 2023).

Dittrich, M., Rathje, J., Manemann, T., & Müller, F. (2022). *Militant accelerationism origins and developments in Germany*. Berlin: Center for Monitoring and Analysis (CeMAS).

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology, 31*(5), 410–436.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society, 21*(5), 729–745.

Earnest, J. (2019). *An open letter.* The Poway Synagogue shooter manifesto, Poway.

Earth Charter International Secretariat (2000). *Earth Charter.* https://earthcharter.org/library/the-earth-charter-in-action-2000-annual-report/ (retrieved 19 June 2023).

Ebner, J. (2017). *The rage: The vicious circle of Islamist and far-right extremism.* London: Bloomsbury Academic.

Ebner, J. (2019). Replatforming unreality: An examination of alt-tech online ecosystems and the dangers these hotbeds of self-reinforcing extremism pose to security, politics and society. *Journal of Design and Science, 6*. https://doi.org/10.21428/7808da6b.e585ddcb.

Ebner, J. (2020). *Going dark: The secret social lives of extremists.* London: Bloomsbury.

Ebner, J. (2022, April). *Why we no longer get along*. Youtube Channel of TEDx Oxford. https://www.ted.com/talks/julia_ebner_why_we_no_longer_get_along_apr_2022 (retrieved 20 January 2023).

Ebner, J., Kavanagh, C., & Whitehouse, H. (2022a). Is there a language of terrorists? A comparative manifesto analysis. *Studies in Conflict and Terrorism*. https://doi.org/10.1080/1057610X.2022.2109244.

Ebner, J., Kavanagh, C., & Whitehouse, H. (2022b). The QAnon security threat: A linguistic fusion-based violence risk assessment. *Perspectives on Terrorism, 16*(6), 62–86.

Ebner, J. (2023). *Going Mainstream: how extremists are taking over. London: Bonnier.*

Ebner, J., & Whitehouse, H. (2023). The identity-extremism nexus: Socio-psychological radicalisation drivers. In J. Busher, L. Malikki, & S. Marsden (Eds.), *Routledge Handbook on Radicalisation and Counter Radicalisation.* Abingdon: Routledge.

el-Husseini, R. (2008). Resistance, Jihad, and martyrdom in contemporary Lebanese Shi'a discourse. *Middle East Journal, 62*(3), 399–414.

Emmerson, B. (2016). *Report of the Special Rapporteur on the promotion of human rights and fundamental freedoms while countering terrorism*. Human Rights Council.

Enders, A. M., Uscinski, J. E., Klofstad, C. A., Wuchty, S., Seelig, M. I., Funchion, J. R., Murthi, M. N., Premaratne, K., & Stoleret, J. (2022). Who supports QAnon? A case study in political extremism. *The Journal of Politics, 84*(3), 1844–1849.

Escobar, A., Hess, D., Licha, I., Sibley, W., Strathern, M., & Sutz, J. (1994). Welcome to cyberia: Notes on the anthropology of cyberculture. *Current Anthropology,* 35(3), 211–231.

European Commission (2022). *The Digital Services Act package. Shaping Europe's digital future.* Brussels: European Union Publication. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package (retrieved 10 June 2023).

Europol. (2021). *European Union terrorism situation and trend report.* The Hague: Europol.

Extinction Rebellion (2018). *Declaration of Rebellion*. https://extinctionrebellion.uk/declaration/ (retrieved 19 June 2023).

Federal Bureau of Investigation [FBI]. (2022). *What we investigate: FBI threat intimidation guide*. Counter-Intelligence. https://www.fbi.gov/investigate/counterintelligence/threat-intimidation-guide#Eletronic-message%20threat

Fein, R. A., & Vossekuil, B. (2000). *Protective intelligence threat assessment investigations: A guide for state and local law enforcement officials*. Washington, DC: U.S. Department of Justice.

Ferguson, J. (2014, 15–18 May). A rightful share: Beyond gift and market in the politics of distribution. Keynote address at The Future with/of Anthropologists, Japanese Society of Cultural Anthropology, Tokyo.

Fielitz, M. & Ahmed, R (2021). I*t's not funny anymore. Far-right extremists' use of humour.* Brussels: European Commission, Radicalisation Awareness Network.

Fielitz, M., Ebner, J., Guhl, J., & Quent, M. (2018). Loving hate: Anti-Muslim extremis*m, radical Islamism and the spiral of polarization.* Jena, London, Berlin: IDZ and Institute for Strategic Dialogue.

Fielitz, M., & Marcks, H. (2019). *Digital fascism: Challenges for the open society in times of social media.* Working Paper Series. Institute for Peace Research and Security Policy at the University of Hamburg and Berkeley Center for Right-Wing Studies.

Fielitz, M., & Thurston, N. (2019). *Post-digital cultures of the far right: Online actions and offline consequences in Europe and the US.* Bielefeld: Transcript.

Fitria, K., & Febrianti, Y. (2020). The interpretation and attitude of body shaming behavior on social media: A digital ethnographic study on Instagram. *Diakom: Jurnal Media dan Komunikasi, 3*(1), 12–25.

Fjordman. (2008). *Defeating Eurabia.* Self-published book.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online consumption. *Public Opinion Quarterly, 80*(S1), 298–320.

Fraise, T. (2017). *Shades of jihad: Variation of military ethics between ISIS and alQaeda*. Paris: Sciences Po Kuwait Program.

Fridays4Future. (2022, 22 October). fridaysforfuture.org (retrieved 22 October 2022).

Gallagher, A., Davey, J., & Hart, M. (2020). *The genesis of a conspiracy theory: Key trends in QAnon activity since 2017*. London: Institute for Strategic Dialogue.

Gardell, M. (2014). Crusader dreams: Oslo 22/7, Islamophobia, and the quest for a monocultural Europe. *Terrorism and Political Violence, 26*, 129–155.

Gerster, L., Kuchta, R., Hammer, D., & Schwieter, C. (2021). *Stützpfeiler Telegraam: Wie Rechtsextreme und Verschwörungsideolog: innen auf Telegram ihre Infrastruktur ausbauen* [*Supporting pillars of Telegram: How right-wing extremists and conspiracy ideologists are expanding their infrastructure on Telegram*]. Berlin: Institute for Strategic Dialogue.

Gibson, W. (1984). *Neuromancer.* New York: Ace.

Gill, P., Horgan, J., & Deckert, P. (2014). Bombing alone: Tracing the motivations and antecedent behaviors of lone-actor terrorists. *Journal of Forensic Science, 59*(2), 425–435.

Giner-Sorolla, R., Leidner, B., & Castano, E. (2011). Dehumanization, demonization, and morality shifting. In M. H. Blaylock (Ed.), *Extremism and the psychology of uncertainty* (pp. 165–182). Hoboken, NJ: Wiley-Blackwell.

Ginsburg, F. (2012). Disability in the digital age. In H. Horst, & D. Miller (Eds.), *Digital anthropology* (pp. 101–126). Oxford: Berg.

Glatt, Z. (2021). We're all told not to put our eggs in one basket: Uncertainty, precarity and cross-platform labor in the online video influencer industry. *International Journal of Communication 16*, 1–19.

Glatt, Z. (2022). Precarity, discrimination and (in)visibility: An ethnography of "the algorithm" in the YouTube influencer industry. In E. Costa, P. G. Lange, N. Haynes, & J. Sinanan (Eds.), *The Routledge Companion to Media Anthropology.* Abingdon: Routledge.

Gómez, A., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., & Swann, W. B. (2011). On the nature of identity fusion: Insights into the construct and a new measure. *Journal of Personality and Social Psychology, 100*, 918–933.

Gøtzsche-Astrup, O. (2018). The time for causal designs: Review and evaluation of empirical support for mechanisms of political radicalisation. *Aggression and Violent Behavior, 39*, 90–99.

Government Offices of Sweden. (2011). *Action plan to safeguard democracy against violence-promoting extremism*. Stockholm: Regeringskansliet.

Gray, J. (2022, 8 August). Sony's a7 IV gets anti-forgery crypto signature technology (for commercial purposes). *Digital Photography Review*.

Griera, M., Gras, J. M., Clot-Garrell, A., & Cazarín, R. (2022). Conspirituality in COVID-19 times: A mixed-method study of anti-vaccine movements in Spain. *Journal for the Academic Study of Religion, 35*(2), 192–217.

Gruenewald, J., Chermak, S., & Freilich, J. D. (2013). Distinguishing "loner" attacks from other domestic extremist violence: A comparison of far-right homicide incident and offender characteristics. *Criminology and Public Policy, 12*(1), 65–91.

Guess, A., Lyons, B., Nyhan, B., & Reifler, J. (2018). *Avoiding the echo chamber about the echo chamber: Why selective exposure to like-minded political news is less prevalent than you think.* Miami, FL: Knight Foundation.

Guhl, J. (2018). Why beliefs always matter, but rarely help us predict jihadist violence. The role of cognitive extremism as a precursor for violent extremism. *Journal for Deradicalization, 14*, 192–217.

Guhl, J., Ebner, J., & Rau, J. (2020). *The online ecosystem of the German far-right*. London/Berlin: Institute for Strategic Dialogue and Bosch Foundation.

Habash, N. Y. (2010). *Introduction to Arabic natural language processing.* San Rafaelm CA: Morgan & Claypool Publishers.

Hafez, M. M. (2007). Martyrdom mythology in Iraq: How Jihadists frame suicide terrorism in videos and biographies. *Terrorism and Political Violence, 19*(1), 95–115.

Halverson, J. (2017). Yusuf al-Qaradawi. Oxford Bibliographies. https://www.oxfordbibliographies.com/view/document/obo-9780195390155/obo-9780195390155-0098.xml (retrieved 1 June 2022).

Han, J., Heo, J., & You, E. (2021). Analysis of metaverse platform as a new play culture: Focusing on Roblox and ZEPETO. International Conference on Human-centered Artificial Intelligence.

Harris, K., Gringart, E., & Drake, D. (2014). Understanding the role of social groups in radicalisation. Australian Security and Intelligence Conference, Edith Cowan University. http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1034&context=asi (retrieved 4 November 2019).

Hartleb, F. (2020). *Lone Wolves: The New Terrorism of Right-Wing Single Actors.* Berlin: Springer.

HateAid (2023). *Betroffenenberatung.* https://hateaid.org (retrieved 10 June 2023).

Hersh, E. D., & Schaffner, B. F. (2013). Targeted campaign appeals and the value of ambiguity. *The Journal of Politics, 75*(2), 520–534.

Hiel, A. V., & Mervielde, I. (2003). The measurement of cognitive complexity and its relationship with political extremism. *Political Psychology, 24*(4), 781–801.

Hine, C. (2015). *Ethnography for the internet: Embedded, embodied and everyday.* London: Bloomsbury Academic.

Hirschberger, G., Ein-Dor, T., Leidner, B., & Saguy, T. (2016). How is existential threat related to intergroup conflict? Introducing the multidimensional existential threat (MET) model. *Frontiers in Psychology, 7,* e1877.

Hitler, A. (1981). *Mein Kampf.* (English translation by J.V. Murphy). London: Hurst and Blackett.

HM Government. (2007). *The definition of terrorism: A report by Lord Carlile of Berriew Q.C. independent reviewer of terrorism legislation*. London: HM Government.

HM Government. (2010). *Channel: Supporting individuals vulnerable to recruitment by violent extremists*. London: HM Government.

HM Government. (2012). *Channel: Vulnerability assessment framework*. London: HM Government.

HM Government. (2015). *Counter-Terrorism and Security Act 2015*. London: HM Government. http://www.legislation.gov.uk/ukpga/2015/6/contents/enacted (retrieved 15 March 2020).

HM Government. (2011, June). *Prevent strategy: Presented to Parliament by the Secretary of State for the Home Department by Command of Her Majesty*. London: HM Government. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/97976/prevent-strategy-review.pdf (retrieved 20 November 2019).
Hoffman, B., War, J., & Shapiro, E. (2020). *Assessing the threat of incel violence*. St Andrew's University.
Hogg, M., & Adelman, J. (2013). Uncertainty–identity theory: Extreme groups, radical behavior, and authoritarian leadership. *Journal of Social Issues, 69*(3), 436–454.
Hogg, M., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory, 16*(1), 7–30.
Holdsworth, R. (2020). Remote education in times of COVID-19: A digital ethnographic study of the 'New Normal' in some parts of the world. *Connect, 246*, 18–24.
Holoyda, B. J. (2022). The QAnon conspiracy theory and the assessment of its believers. *Journal of the American Academy of Psychiatry and the Law, 50*(1), 124–135.
Holt, J., & Rizzuto, M. (2021). *QAnon's hallmark catchphrases evaporating from the mainstream internet*. Washington, DC: Atlantic Council's Digital Forensic Research Lab (DFRLab).
Home Office. (2011). *Full equality impact assessment*. London: Office for Security and Counter-Terrorism.
Hook, S. (1973). Myth and fact in the Marxist theory of revolution and violence. *Journal of the History of Ideas, 34*(2), 271–280.
Horst, H. A., & Miller, D. (2006). *The cell phone: An anthropology of communication.* Oxford: Berg.
Horst, H. A., & Miller, D. (2012). *Digital anthropology.* London: Berg.
Houtman, G., & Zeitlyn, D. (1996). Information technology and anthropology. *Anthropology Today, 12*(3), 1–3.
Huang, C.-R., & Chen, K.-J. (1996). Segmentation standard for Chinese natural language processing. *COLING '96 Proceedings of the 16th Conference on Computational linguistics Vol 2*, pp. 1045–1048.
Ihler, B. (2022, 15 May). *On the terrorist attack in Buffalo*. Khalifa Ihler Institute.
Ingram, H. J. (2016). *A "linkage-based" approach to combating militant Islamist propaganda: A two-tiered framework for practitioners*. The Hague: International Centre for Counter-Terrorism.
Institute for Economics & Peace. (2019). *Global Terrorism Index 2019: Measuring the impact of terrorism*. Sydney: Institute for Economics & Peace; National Consortium for the Study of Terrorism and Responses to Terrorism.
Institute for Strategic Dialogue [ISD]. (2019a). *The battle for Bavaria: Online information campaigns in the 2018 Bavarian state election*. London: Institute for Strategic Dialogue.
Institute for Strategic Dialogue [ISD]. (2019b). *Strong cities network*. London: Institute for Strategic Dialogue. https://strongcitiesnetwork.org/en/incorporating-new-approaches-our-approach-to-evaluating-p-cve-interventions/ (retrieved 24 November 2019).

Institute for Strategic Dialogue [ISD]. (2022). *Extremism definition*. London: Institute for Strategic Dialogue, https://www.isdglobal.org/extremism/ (retrieved 17 October 2022).
Institute for Strategic Dialogue [ISD] (2023). *Digital Citizenship Education Programming Toolkit*. London: Institute for Strategic Dialogue. https://www.isdglobal.org/isd-publications/digital-citizenship-education-programming-toolkit/ (retrieved 17 October 2022).
Islam, G. (2014). Social identity theory. In T. Teo (Ed.), *Encyclopedia of critical psychology.* New York, NY: Springer.
Jamal, A. A., Keohane, R. O., Romney, D., & Tingley, D. (2015). Anti-Americanism and anti-interventionism in Arabic Twitter discourses. *Perspectives on Politics, 13*(1), 55–73.
Jamieson, K. H., & Cappella, J. N. (2010). *Echo chamber: Rush Limbaugh and the conservative media establishment.* Oxford: Oxford University Press.
Jensen, M., & Kane, S. (2021). *QAnon offenders in the United States*. College Park, MD: START.
Jeshion, R. (2013). Slurs and stereotypes. *Analytic Philosophy, 54*(3), 314–329.
Jiménez, J., Gómez, A., Buhrmester, M. D., Alexandra Vázquez, H. W., & Swann, W. B. (2016). The dynamic identity fusion index: A new continuous measure of identity fusion for web-based questionnaires. *Social Sciences Computer Review, 34*(2), 215–228.
Johnson, L. (2022). Adapt and adjust: Doing UK-based ethnographic fieldwork during the Covid-19 pandemic. *Cross-Cultural Studies in Health and Illness, 41*(8), 763–777.
Jong, J., Whitehouse, H., Kavanagh, C., & Lane, J. (2015). Shared negative experiences lead to identity fusion via personal reflection. *PLos ONE, 10*(12), 1–12.
Jurgenson, N. (2011, 24 February). *Digital dualism versus augmented reality*. Cyborgology. https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality (retrieved 12 January 2023).
Kaati, L., Shrestha, A., & Cohen, K. (2016). *Linguistic analysis of lone offender manifestos*. Uppsala: Swedish Defence Research Agency.
Kalla, J. L., & Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review, 112*(1), 148–166.
Kavanagh, C., Kapitány, R., Putra, I. E., & Whitehouse, H. (2020). Exploring the pathways between transformative group experiences and identity. *Frontiers in Psychology, 11.* https://doi.org/10.3389/fpsyg.2020.01172.
Kavanagh, C., Wibisono, S., Kapitány, R., Yustisia, W., Putra, I. E., Rufaedah, A., & Whitehouse, H. (2019). Exploring the role of identity fusion and group identification in predicting parochialism amongst Indonesian Islamic groups. *PsyArXiv*. https://doi.org/10.31234/osf.io/e8ytr.
Keady, J. (2021). Querdenker, Querfront, and QAnon: On the German far-right and its American occupation. *EuropeNow.* https://www.europenowjournal.org/2021/07/25/querdenker-querfront-and-qanon-on-the-german-far-right-and-its-american-occupation/.
Keen, F., Crawford, B., & Suarez-Tangil, G. (2020). *Memetic Irony and the promotion of violence within chan cultures*. London: CREST.

Kenyon, J., Binder, J., & Baker-Beall, C. (2021). *Exploring the role of the Internet in radicalisation and offending of convicted extremists*. London: Ministry of Justice.
Kepel, G. (2003). *Jihad: The trail of political Islam.* Cambridge, MA: Harvard University Press.
Kepel, G. (2015). *Terreur dans l'Hexagone: Genèse du djihad français* [*Terror in France: Genesis of French jihad*]. Paris: Gallimar.
Kershaw, I. (1999). *Hitler, 1889–1936: Hubris.* New York, NY: W.W. Norton and Company.
Khalil, J., Horgan, J.G. & Zeuthen, M. (2019). The Attitudes-Behaviors Corrective (ABC) Model of Violent Extremism. *Terrorism and Political Violence 34(3),* 1-26.
King Jr., M. (1963, 28 August). *I Have a Dream*. Washington, DC.
Kirschke-Schwartz, M., & Clarke, C. (2021). *How will violent extremists use technology in the future?* London: Global Network on Extremism and Technology (GNET).
Klein, A. (2019). From Twitter to Charlottesville: Analyzing the fighting words between the Alt-Right and Antifa. *International Journal of Communication, 13,* e10076.
Klein, C., Clutton, P., & Dunn, A. (2019). Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit's conspiracy theory forum. *PLoS ONE, 14*(11), e0225098.
Klein, O. (2020, 15 July). *The Effects of censoring the far-right online*. VOX-Pol. https://www.voxpol.eu/the-effects-of-censoring-the-far-right-online/ (retrieved 10 June 2023).
Koehler, D. (2019). *Violence and terrorism from the far-right: Policy options to counter an elusive threat*. The Hague: International Center for Counter Terrorism.
Koehler, D. & Horgan, J. (2016). *Understanding Deradicalization: Methods, Tools and Programs for Countering Violent Extremism.* Abingdon: Routledge.
Kreps, S. (2021). *Democratizing harm: Artificial intelligence in the hands of nonstate actors*. Washington, DC: Brookings.
Kriner, M., Conroy, M., & Ashwal, Y. (2021). *Understanding accelerationist narratives: 'There is no political solution'*. London: GNET/King's College.
Krotzek, L. J. (2019). Inside the voter's mind: The effect of psychometric microtargeting on feelings toward and propensity to vote for a candidate. *International Journal of Communication, 13*, 3609–3629.
Kruglanski, A., & Fishman, S. (2009). Psychological factors in terrorism and counterterrorism: Individual, group, and organizational levels of analysis. *Social Issues and Policy Review, 3*(1), 1–44.
Kruglanski, A. W., Gunaratna, R., Ellenberg, M., & Speckhard, A. (2020). Terrorism in time of the pandemic: Exploiting mayhem. *Global Security: Health, Science and Policy, 5*(1), 121–132.
Kruglanski, A., Jasko, K., & Webber, D. (2018). The making of violent extremists. *Review of General Psychology, 22*(1). https://doi.org/10.1037/gpr0000144.
Kundnani, A. (2012a). Radicalization: The journey of a concept. *Race and Class, 54*(2), 3–25.
Kundnani, A. (2012b). *The Anti-Islamist: Anders Behring Breivik's Manifesto.* International Centre for Counter Terrorism.
Kundnani, A. (2015). *A decade lost: Rethinking radicalisation and extremism.* London: Claystone.

Kupper, J., Christensen, T. K., Wing, D., Hurt, M., Schumacher, M., & Meloy, R. (2022). The contagion and copycat effect in transnational far-right terrorism: An analysis of language evidence. *Perspectives on Terrorism, 1*(4), 4–26.

Lakhani, S. (2022). *The gamification of violent extremism: An empirical exploration of the Christchurch attack*. London: Global Network on Extremism and Technology (GNET).

Lamphere-Englund, G., & White, D. J. (2022). *The Buffalo attack and the gamification of violence*. London: RUSI.

Lankford, A. (2011). Could suicide terrorists actually be suicidal? *Studies in Conflict & Terrorism, 34*(4), 337–366.

Lankford, A. (2014). Précis of the myth of martyrdom: What really drives suicide bombers, rampage shooters, and other self-destructive killers. *Behavioral and Brain Sciences, 37*, 351–393.

Lankford, A., & Gillespie, K. (2011). Rehabilitating terrorists through counter-indoctrination: Lessons learned from the Saudi-Arabian programme. *International Criminal Justice Review, 21*(2), 118–133.

Leach, C., Zomeren, M. V., S. Zebel, M. V., Pennekamp, S., Doosje, B., Ouwekerk, J., & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology, 91*(1), 144–165.

Lee, B., & Knott, K. (2020). More grist to the mill? Reciprocal radicalisation and reactions to terrorism in the far-right digital milieu. *Perspectives on Terrorism, 14*(3), 98–115.

Leidig, E., & van Mieghem, C. (2021). *The US national strategy on countering domestic terrorism as a model for the EU*. The Hague: International Centre for Counter-Terrorism.

Liao, S. (2018, 28 February). Discord shuts down more neo-Nazi, alt-right servers. *The Verge*.

Lickel, B., Miller, N., Stenstrom, D., Denson, T., & Schmader, T. (2006). Vicarious retribution: The role of collective blaming in intergroup aggression. *Personality and Social Psychology Review, 10*(4), 372–390.

Lifton, R. (1988). *The Nazi doctors: Medical killing and the psychology of genocide.* New York, NY: Basic Books.

Lister, M., & Jarvis, L. (2013). Disconnection and resistance: Anti-terrorism and citizenship in the UK. *Citizenship Studies, 17*(6–7), 756–769.

Louis, W. R., McGarty, C., Thomas, E. F., Amiot, C. E., & Moghaddam, F. M. (2018). The power of norms to sway fused group members. *Behavioral and Brain Sciences, 41*, 1–62.

Louis, W. R., Taylor, D. M., & Douglas, R. L. (2005). Normative influence and rational conflict decisions: Group norms and cost-benefit analyses for intergroup behavior. *Group Processes & Intergroup Relations, 8*(4), 355–374.

Lupton, D. (2014). *Digital sociology.* London: Routledge.

Lygre, R., Eid, J., Larsson, G., & Ranstorp, M. (2011). Terrorism as a process: A critical review of Moghaddam's "staircase to terrorism". *Scandinavian Journal of Psychology, 52*(6), 609–616.

Lynch, O., & Joyce, C. (2018). Functions of collective victimhood: Political violence and the case of the Troubles in Northern Ireland. *International Review of Victimology, 24*(2), 183–197.
Macfarlane, A., & Turin, M. (2020). *Digital Himalaya*. http://www.digitalhimalaya.com/ (retrieved 21 September 2022).
MacKee, F. (2016). Social media in gay London: Tinder as an alternative to hook-up apps. *Social Media & Society*, 1–10.
Macklin, G. (2019). The Christchurch attacks: Livestream terror in the viral video age. *CTC Sentinel, 12*(6). https://ctc.westpoint.edu/christchurch-attacks-livestream-terror-viral-video-age/.
Mael, F., & Ashforth, B. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior, 13*, 103–123.
Mael, F. A., & Tetrick, L. E. (1992). Identifying organizational identification. *Educational and Psychological Measurement, 52*, 813–824.
Maizland, L. (2019). *China's repression of Uighurs in Xinjiang*. New York, NY: Council on Foreign Relations.
Martin-Borregon, D., Aiello, L. M., Grabowicz, P., Jaimes, A., & Baeza-Yates, R. (2014). Characterization of online groups along space, time, and social dimensions. *EPG Data Science 3*(8). https://doi.org/10.1140/epjds/s13688-014-0008-y
Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. New York, NY: Data & Society.
Marx, K., & Engels, F. (2004). *Manifesto of the Communist Party*. English edition by Andy Blunden.
Mashuri, A., & Zaduqisti, E. (2015). The effect of intergroup threat and social identity salience on the belief in conspiracy theories over terrorism in Indonesia: Collective angst as a mediator. *International Journal of Psychological Research, 8*(1), 24–35.
Matfess, H., & Margolin, D. (2022). *The women of January 6th*. Washington, DC: Program on Extremism at George Washington University.
Mattheis, A. A. (2022). *Atomwaffen Division & its affiliates on Telegram: Variations, practices, & interconnections*. RESOLVE Network.
McCauley, C. (1972). Extremity shifts, risk shifts and attitude shifts after group discussion. *European Journal of Social Psychology, 2*(4), 417–436.
McCauley, C., & Moskalensko, S. (2008). Mechanisms of political radicalization: Pathways to terrorism. *Terrorism and Political Violence, 20*(3), 415–433.
McCauley, C., & Moskalenko, S. (2017). Understanding political radicalization: The two-pyramids model. *American Psychologist, 72*(3), 205–216.
McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., & Landwehr, J. (2017). Using social media to monitor mental health discussions − Evidence from Twitter. *Journal of the American Medical Informatics Association, 24*(3), 496–502.
McManus, J. (2018). Modern enchantments: Media, fandom and distraction amongst Turkish football supporters. *Ethnos, 83*(4), 762–781.
Meleagrou-Hitchens, A., & Kaderbhai, N. (2018). *Research perspectives on online radicalisation: A literature review*. London: International Centre for the Study of Radicalisation, King's College.

Meta (2023). *Hate speech policy details.* Transparency Center. Facebook Community Standards. https://transparency.fb.com/en-gb/policies/community-standards/hate-speech (retrieved 10 June 2023).

Metropolitan Police Service. (2008). *Equality impact assessment.* London: HM Government.

Mies, M. & Shiva, V. (1993). *Ecofeminism.* Halifax: Fernwood Publications.

Miller, C. (2016). *From Brussels to Brexit: Islamophobia, xenophobia, racism and reports of hateful incidents on Twitter*. London: Demos.

Miller, D. (2011). *Tales from Facebook.* Cambridge: Polity Press.

Miller, D. (2016). *Social media in an English village: Or how to keep people at just the right distance.* London: UCL Press.

Miller, D., & Slater, D. (2000). *The internet: An ethnographic approach.* Oxford: Berg.

Miller, J., Bland, N., & Quinton, P. (2000). *The impacts of stops and searches on crime and the community*. Police Research Series, Paper 127. London: Home Office.

Miller-Idriss, C. (2021). *White supremacist extremism and the far right in the U.S.* Farmington Hills, MI: Gale Primary Sources/Cengage.

Ministry of Justice. (2009). *Statistics on race and the criminal justice system 2007/8. London: A ministry of justice publication under section 95 of the Criminal Justice Act 1991*. London: Ministry of Justice.

Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist, 60*(2), 161–169.

Monteil, C. (1997). Simone de Beauvoir and the women's movement in France: An eye-witness account. *Simone de Beauvoir Studies, 14*, 6–12.

Moskalenko, S., & McCauley, C. (2021). QAnon: Radical opinion versus radical action. *Perspectives on Terrorism, 15*(2), 142–146.

Munn, L. (2019). Alt-right pipeline: Individual journeys to extremism online. *First Monday, 24*(6). https://doi.org/10.5210/fm.v24i6.10108

Mussiraliyeva, S., Bolatbek, M., Omarov, B., Medetbek, Z., Baispay, G., & Ospanov, R. (2020). On detecting online radicalization and extremism using natural language processing. *21st International Arab Conference on Information Technology (ACIT)*, pp. 28–30.

Nardi, B. (2010). *My life as a night elf priest: An anthropological account of World of Warcraft.* Ann Arbor, MI: University of Michigan Press.

National Commission on Terrorist Attacks upon the United States. (2004). *The 9/11 Commission report: Final report of the National Commission on Terrorist Attacks upon the United States*. Washington, DC: National Commission on Terrorist Attacks upon the United States.

National Counterterrorism Center [NCTC], Federal Bureau of Investigation [FBI], & U.S. Department of Homeland Security [DHS]. (2021). *U.S. violent extremist indicators* (2021 edition). Washington, DC: National Counterterrorism Center.

Nemeth, S. C., Mauslein, J. A., & Stapley, C. (2014). The primacy of the local: Identifying terrorist hot spots using geographic information systems. *The Journal of Politics, 76*(2), 304–317.

Neumann, P. R. (2013). Options and strategies for countering online radicalization in the United States. *Studies in Conflict and Terrorism, 36*(6), 431–459.

Neumann, P. R. (2015). *Victim, perpetrators, assets: The narratives of Islamic State defectors*. London: The International Centre for the Study of Radicalisation and Political Violence.
Neumann, P. R. (2017). *Countering violent extremism and radicalisation that lead to terrorism: Ideas, recommendations, and good practices from the OSCE region.* Vienna: Organization for Security and Cooperation in Europe.
Newhouse, A. (2021). The threat is the network: The multi-node structure of neo-fascist accelerationism. *CTC Sentinel, 14*(5). https://ctc.westpoint.edu/the-threat-is-the-network-the-multi-node-structure-of-neo-fascist-accelerationism/
Newson, M. (2019). Football, fan violence, and identity fusion. *International Review for the Sociology of Sport, 54*(4), 431–444.
Newson, M., Bortolini, T., Buhrmester, M., Silva, S. R., Nicássio, J., Aquino, Q., & Whitehouse, H. (2018). Brazil's football warriors: Social bonding and inter-group violence. *Evolution and Human Behavior, 39*(6), 675–683.
Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs*. Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, pp. 93–98.
Norwegian Police Security Service [PST]. (2022). *National threat assessment for 2022*. https://pst.no/alle-artikler/trusselvurderinger/ntv-2022/
Nossiter, A. (2016, 12 July). 'That ignoramus': 2 French scholars of radical Islam turn bitter rivals. *New York Times*. https://www.nytimes.com/2016/07/13/world/europe/france-radical-islam.html (retrieved 16 November 2019).
NPR/Ipsos. (2020, 20 December). Even if it's 'bonkers,' poll finds many believe QAnon and other conspiracy theories. NPR. https://www.npr.org/2020/12/30/951095644/even-if-its-bonkers-poll-finds-many-believe-qanon-andother-conspiracy-theories (retrieved 19 November 2022).
O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1–13.
Office of the Director of National Intelligence. (2021). *Domestic violent extremism poses heightened threat in 2021*. Washington, DC: Office of the Director of National Intelligence.
O'Hara, K., & Stevens, D. (2015). Echo chambers and online radicalism: Assessing the internet's complicity in violent extremism. *Policy and Internet, 7*(4), 401–422.
Organization for Security and Cooperation in Europe [OSCE]/Office for Democratic Institutions and Human Rights [ODIHR]. (2018). *Guidelines for addressing the threats and challenges of "foreign terrorist fighters" within a human rights framework*. Poland: OSCE Office for Democratic Institutions and Human Rights (ODIHR).
Ovide, S. (2021, 23 April). When is online nastiness illegal? *The New York Times*.
Pape, R. (2021). *American political violence: Why we cannot afford to ignore the American insurrectionist movement.* Chicago, IL: Chicago Project on Security & Threat (CPOST).
Pape, R. (2022a, 6 January). The Jan. 6 insurrectionists aren't who you think they are. *Foreign Policy*. https://foreignpolicy.com/2022/01/06/trump-capitol-

insurrection-january-6-insurrectionists-great-replacement-white-nationalism/ (retrieved 2 February 2023).
Pape, R. (2022b). *The Jan 6th hearings are reducing the violent threat to democracy, according to latest survey*. Chicago, IL: Chicago Project on Security and Threats (CPOST).
European Parliament. (2014). *Preventing and countering youth radicalisation in the EU.* Brussels: EU Directorate General for Internal Policies.
Parmar, A. (2011). Stop and search in London: Counter-terrorist or counter-productive? *Policing and Society, 21*(4), 375–377.
Pearson, E. (2018). Online as the new frontline: Affect, gender, and ISIS-take-down on social media. *Studies in Conflict & Terrorism 31* (11), 850-874.
Pemberton, A., & Aarten, P. (2018). Narrative in the study of victimological processes in terrorism and political violence: An initial exploration. *The Complexity of Terrorism - Victims, Perpetrators and Radicalisation, 24*(2), 183–197.
Pennebaker, J., & Chung, C. K. (2008). Computerized text analysis of Al-Qaeda transcripts. In K. Krippendorf, & M. A. Bock (Eds.), *The content analysis reader* (pp. 453–465). Newbury Park, CA: Sage.
Polarization and Extremism Research and Innovation Lab [PERIL] and Network Contagion Research Institute. (2020). *The QAnon conspiracy: Destroying families, dividing communities, undermining democracy*. Polarization and Extremism Research and Innovation Lab and Network Contagion Research Institute.
Poscheschnik, G. & Crepaldi, G. (2022). Only chance and circumstances? Or, how Freudian are Freudian slips? A review of research literature concerning parapraxes. *Psychoanalytic Psychology, 39*(3), 189-197.
Peters, M. A., & Besley, T. (2017). White supremacism: The tragedy of Charlottesville. *Educational Philosophy and Theory, 49*(14), 1309–1312.
Pink, S., Horst, H., Postrill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2016). *Digital ethnography: Principles and practice.* London: Sage.
Pisou, D., & Ponesh, S. (2016). *Youth, families and communities*. Brussels: Radicalisation Awareness Network (RAN).
Postill, J. (2011). *Localizing the internet: An anthropological account.* Oxford: Berghahn.
Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology, 52*, 597–617.
Powell, L. (2016). Counter-productive counter-terrorism: How is the dysfunctional discourse of Prevent failing to restrain radicalisation? *Journal for Deradicalisatio, 8*, e6.
Python, A., Bender, A., Nandi, A. K., Hancock, P. A., Arambepola, R., Brandsch, J., & Lucas, T. C. (2021). Predicting non-state terrorism worldwide. *Science Advances, 7*(31). https:/doi.org/10.1126/sciadv.abg4778.
Python, A., Illian, J., Jones-Todd, C., & Blangiardo, M. (2018). A Bayesian approach to modelling subnational spatial dynamics of worldwide non-state terrorism, 2010–2016. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 182*(1), 323–344.

Qi, Y. (2020). The language of terror: Exploring speech acts in official English-language ISIS videos, 2014–2017. *Small Wars and Insurgencies, 31*(6), 1196–1241.
Quaranto, A., & Stanley, J. (2021). Propaganda. In J. Khoo, & R. K. Sterken (Eds.), *The Routledge Handbook of Social and Political Philosophy of Language.* Abingdon: Routledge.
Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on Facebook. Harvard Law School Discussion Paper 877.
Qutb, S. (2006). *Milestones.* Delhi: Islamic Book Service.
Radicalisation Awareness Network [RAN]. (2021). *The gamification of violent extremism & lessons for P/CVE.* Luxembourg: Publications Office of the European Union.
Ranstorp, M. (2010). *Understanding violent radicalisation: Terrorist and jihadist movements in Europe.* London: Routledge.
Reddy, M., & Hanumanthappa, M. (2014). Semantical and syntactical analysis of NLP. *International Journal of Computer Science and Information Technologies, 5*(3), 3236–3238.
Reicher, S., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology, 6*, 161–198.
Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier.* Reading, MA: Addison-Wesley.
Richter, E. D., Markus, D. K., & Tait, C. (2018). Incitement, genocide, genocidal terror, and the upstream role of indoctrination: Can epidemiologic models predict and prevent? *Public Health Review, 39*(30). https://doi.org/10.1186/s40985-018-0106-7
Roberts, H., & Floridi, L. (2021). *The EU and the US: Two different approaches to AI governance*. Oxford: Oxford Internet Institute.
Roberts, S. R. (2018). The biopolitics of China's "war on terror" and the exclusion of the Uyghurs. *Critical Asian Studies, 50*(2), 232–258.
Rodger, E. (2014). *My twisted world: The story of Elliot Rodger*. Terrorist manifesto, Isla Vista.
Roof, D. (2015). *Manifesto of the Charleston shooting*. The Last Rhodesian website.
Rosenberg, M. (1987). Depersonalisation: the loss of personal identity. In T. Honess, & K. Yardley (Eds.), *Self and identity: Perspectives across the lifespan* (pp. 193–206). Abingdon: Routledge.
Roy, O. (2015, 19 November). What is the driving force behind jihadist terrorism? – A scientific perspective on the causes/circumstances of joining the scene. Speech at BKA Autumn Conference: International Terrorism: How can prevention and repression keep pace?. Berlin: Bundeskriminalamt. https://life.eui.eu/wp-content/uploads/sites/7/2015/11/OLIVIER-ROY-what-is-a-radical-islamist.pdf (retrieved 19 November 2019).
Roy, O. (2017). *Jihad and death: The global appeal of Islamic State.* London: Hurst.
Rubin, O., Bruggeman, L., & Steakin, W. (2021, 20 January). QAnon emerges as recurring theme of criminal cases tied to US Capitol siege. *ABC News*.
Runge, K. K., Yeo, S. K., Cacciatore, M., Scheufele, D. A., Brossard, D., Xenos, M., Anderson, A., Choi, D.-H., Kim, J., Li, N., Liang, X., Stubbings, M., & Su, L. Y.-F. (2013). Tweeting nano: How public discourses about nanotechnology develop in social media environments. *Journal of Nanoparticle Research, 15*(1), 1–11.

Sage Research Methods. (2016). *Digital social research: Centre for the Analysis of Social Media (CASM)*. https://methods.sagepub.com/video/digital-social-research-centre-for-the-analysis-of-social-media-casm (retrieved 3 December 2019).
Sageman, M. (2008). A strategy for fighting international Islamist terrorists. *The Annals of the American Academy of Political and Social Science, 618*, 223–231.
Salazar, J. F., Pink, S., Irving, A., & Sjöberg, J. (2017). *Anthropologies and futures: Researching emerging and uncertain worlds.* London: Bloomsbury.
Saltman, E. (2020). *Countering terrorism and violent extremism at Facebook: Technology, expertise and partnerships.* New Delhi: Observer Research Foundation.
Schiano, C. (2017a). *Charlottesville violence planned over Discord servers: Unicorn riot reports*. Unicorn Riot.
Schiano, C. (2017b). *Anticom Discord server shows months of neo-Nazi incitement*. Unicorn Riot.
Schlegel, L. (2021). Connecting, competing, and trolling: “User types” in digital gamified radicalization processes. *Perspectives on Terrorism, 15*(4), 54–64.
Schroeder, R. (2011). *Being there together: Social interaction in shared virtual environments.* Oxford: Oxford University Press.
Schwarz-Friesel, M. (2015). Using language as a weapon: Verbal manifestations of contemporary anti-Semitism. In R. Schulze, & H. Pishwa (Eds.), *The exercise of power in communication* (pp. 161–183). London: Palgrave Macmillan.
Sessions, G. (1995). *Deep Ecology for the Twenty-First Century*. Boulder, CO: Shambala.
Shaw, A., Khalifa, N., & Hassan, T. (2022). The psychiatric aspects of terrorism: Prevention and rehabilitation. *International Journal of Risk and Recovery, 5*(2), 31–36.
Sheehan, I. (2014). Are suicide terrorists suicidal? A critical assessment of the evidence. *Innovations in Clinical Neuroscience, 11*(9–10), 81–92.
Silber, M. D., & Bhatt, A. (2007). *Radicalization in the West: The homegrown threat.* New York, NY: NYPD Intelligence New York Division.
Silke, A. (2019). *Routledge handbook of terrorism and counterterrorism.* Abingdon: Routledge.
Singer-Emery, J., & Bray, R. (2020). *The Iron March data dump provides a window into how White supremacists communicate and recruit*. New York, NY: Lawfare.
Smedt, T. D., Pauw, G. D., & Ostaeyen, P. V. (2018). *Automatic detection of online jihadist hate speech*. Computational Linguistics & Psycholinguistics Report Series, CTRS-007*.*
Smirnova, J., & Arcostanzo, F. (2022). *Russia-Ukraine war sparks influx of disinformation in German-language conspiracy groups*. London: Institute for Strategic Dialogue.
Smith, A., Suedfeld, P., Conway, L. G., & Winter, D. G. (2008). The language of violence: Distinguishing terrorist from nonterrorist groups by thematic content analysis. *Dynamics of Asymmetric Conflict, 1*(2), 142–163.
Smith, B. L., Snow, D. A., Fitzpatrick, K., Damphousse, K. R., Roberts, P., Tan, A., Brooks, A., & Klein, B. (2016). *Identity and framing theory, precursor activity, and the radicalization process*. Washington, DC: U.S. Department of Justice.

Speckhard, A., Shajkovci, A., & Yayla, A. (2017). Following a military defeat of ISIS in Syria and Iraq: What happens next after the military victory and the return of foreign fighters? *Journal of Terrorism Research, 8*(1), 81–89.
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,* 89-121.
Spini, D., Elcheroth, G., & Fasel, R. (2008). The impact of group norms and generalization of risks across groups on judgments of war behavior. *Political Psychology, 29*(6), 920–941.
Stanford Center for International Security and Cooperation. (2022). *Mapping the global far-right*. Stanford, CA: Stanford Center for International Security and Cooperation.
Sternberg, R. (2003). A duplex theory of hate: Development and application to terrorism, massacres, and genocide. *Review of General Psychology, 7*(3). https://doi.org/10.1037/1089-2680.7.3.299
Stone, A. (1991). Will the real body please stand up? Boundary stories about virtual cultures. In M. Benedikt (Ed.), *Cyberspace: The First Steps*. Cambridge: The MIT Press.
Strindberg, A. (2020). *Social identity theory and the study of terrorism and violent extremism*. Stockholm: Swedish Defence Research Agency.
Sugiura, L. (2021). *The Incel rebellion: The rise of the manosphere and the virtual war against women.* Bingley: Emerald Publishing Limited.
Suler, J. (2014). The online disinhibition effect. *Cyber Psychology & Behaviour, 7*(3), 321–326.
Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
Swann, W. B., & Buhrmester, M. D. (2015). Identity fusion. *Current Directions in Psychological Science, 24*(1), 52–57.
Swann, W., Buhrmester, M., Gómez, A., Jetten, J., Bastian, B., Vázquez, A., Ariyanto, A., Besta, T., Christ, O., Cui, L., Finchilescu, G., González, R., Goto, N., Hornsey, M., Sharma, S., Susianto, H., & Zhang, A. (2014). What makes a group worth dying for? Identity fusion fosters perception of familial ties, promoting self-sacrifice. *Journal of Personality and Social Psychology, 106*, 912–926.
Swann, W., Gómez, A., Dovidio, J., Hart, S., & Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science, 2*(8), 1176–1183.
Swann, W. B., Gómez, A., Seyle, D., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology, 96*, 995–1011.
Swann, W., Jetten, J., Gómez, A., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review, 119*, 441–456.
Tajfel, H., & Turner, J. (1979). The social identity theory of intergroup behavior. In S. Worchel, & W. G. Austin (Eds.), *The social psychology of intergroup relations.* Chicago, IL: Nelson-Hall Publishers.

Tajfel, H., & Turner, J. (1997). An integrative theory of intergroup conflict. In W. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey: Brooks Cole Publishing.

Tarrant, B. (2019). *The great replacement: Towards a new society*. Terrorist manifesto, Christchurch.

Taylor, G., Frederiksen, R., Vane, R., & Waltz, E. (2004). Agent-based simulation of geo-political conflict. *IAAI'04: Proceedings of the 16th Conference on Innovative Applications of Artificial Intelligence*, pp. 884–891.

Technau, B. (2018). Going beyond hate speech: The pragmatics of ethnic slur terms. *Lodz Papers in Pragmatics, 14*(1), 25–43.

Thomas, E. F., & Louis, W. R. (2014). When will collective action be effective? Violent and non-violent protests differentially influence perceptions of legitimacy and efficacy amongst supporters. *Personality and Social Psychology Bulletin, 40*, 263–276.

Thunberg, G. (2020, January). *Our house is still on fire*. Davos: World Economic Forum.

Tomasello, M. (2014). *The new psychology of language: Cognitive and functional approaches to language structure.* New York, NY: Psychology Press.

Torregrosa, J., Bello-Orgaz, G., Martínez-Cámara, E., Ser, J. D., & Camacho, D. (2022). A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-021-03658-z.

Tossell, C., Chad, C., Gómez, A., Ewart, J., Visser, J. D., Vázquez, A., Donadio, B. T., Metcalfe, A., Rogan, C., Davis, R., & Atran, S. (2022). Spiritual over physical formidability determines willingness to fight and sacrifice through loyalty in cross-cultural populations. *Proceedings of the National Academy of Sciences, 119*(6), e2113076119.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization and political disinformation: A review of the scientific literature*. Menlo Park, CA: The Hewlett Foundation.

Tufekci, Z (2018, 10 March). YouTube, the great radicalizer. *The New York Times 10*(3). https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html (retrieved 10 June 2023).

Turkle, S. (1995). *Life on the screen: Identity in the age of the internet.* New York, NY: Simon & Schuster.

UK Parliament (2023). *Online Safety Bill. Government Bill originated in the House of Commons, Sessions 2021-22, 2022-23.* London: HM Government. https://bills.parliament.uk/bills/3137 (retrieved 10 June 2023).

United Nations [UN] (2018, 26 June). *General Assembly unanimously adopts resolution calling for strong coordinated action by member states to tackle terrorism, violent extremism worldwide*. *General Assembly, 72nd Session, GA/12035*. New York, NY: United Nations. https://www.un.org/press/en/2018/ga12035.doc.htm (retrieved 30 January 2023).

United Nations Counter-Terrorism Committee Executive Directorate [UNCTED]. (2016). *Implementation of Security Council resolution 2178 (2014), para. 158(b), op. cit., note 62.* New York, NY: United Nations.

United Nations Development Programme [UNDP]. (2016). *Preventing violent extremism through promoting inclusive development, tolerance and respect for diversity.* Bratislava: United Nations Development Programme (UNDP).

United Nations Office on Drugs and Crime [UNODC]. (2023, 11 March). *Defining terrorism*. E4J University Module Series: Counter-Terrorism, Module 4: Criminal Justice Responses to Terrorism. https://www.unodc.org/e4j/en/terrorism/module-4/key-issues/defining-terrorism.html (retrieved 11 March 2023).

United Nations Security Council. (2019, 20 May). *Security Council members call for updating sanctions lists to ensure constant vigilance against threat of terrorist attacks, 8528th Meeting, SC/13814*. https://www.un.org/press/en/2019/sc13814.doc.htm (retrieved 21 November 2019).

University of Cambridge. (2020). *The Thak Archive*. http://www.digitalhimalaya.com/collections/thakarchive/

University of Groningen. (2013, 16 April). Order, justice and extremism: Martin Luther King, Jr and “Letter from Birmingham Jail” 50 years on. https://www.rug.nl/research/centre-for-religious-studies/research-centres/centre-religion-conflict-globalization/blog/order-justice-and-extremism-martin-luther-king-16-04-2013?lang=en (retrieved 14 March 2023).

University of Oxford. (2020a). *Safeguarding ‘at risk’ adults and children*. https://hr.admin.ox.ac.uk/safeguarding-at-risk-adults-and-children (retrieved 15 January 2021).

University of Oxford. (2020b). *Approved procedures*. https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/ap (retrieved 15 January 2021).

Upchurch, H. (2021). The Iron March forum and the evolution of the ‘skull mask’ neo-fascist network. *CTC Sentinel, 14*(10), 27–37.

U.S. Department of Homeland Security [DHS]. (2018). *Digital forum on terrorism prevention: A report on trends and insights*. Washington, DC: U.S. Department of Homeland Security.

U.S. Department of Justice. (2014). *Violent Extremism. Awareness Brief.* Washington, DC: Office of Community Oriented Policing Services.

U.S. Government Publishing Office. (2020). *Threats to the homeland*. *Hearing before the Committee on Homeland Security and Governmental Affairs, 116th Hearing, 2nd Session.* Washington, DC: United States Senate.

Vaes, J., Leyens, J.-P., Paladino, M. P., & Miranda, M. P. (2012). We are human, they are not: Driving forces behind outgroup dehumanisation and the humanisation of the ingroup. *European Review of Social Psychology, 2*(1), 64–106.

Vázquez, A., Ordoñana, J. R., Whitehouse, H., & Gómez, Á. (2019). Why die for my sibling? The positive association between identity fusion and imagined loss with endorsement of self-sacrifice. *International Journal of Social Psychology, 34*(3), 413–438.

Vegt, I. V., Mozes, M., Gill, P., & Kleinberg, B. (2020). Online influence, offline violence: Language use on YouTube surrounding the ‘Unite the Right’ rally. *Journal of Computational Social Science, 4*, 333–354.

Veilleux-Lepage, Y. (2021). *CTRL, HATE, PRINT: Terrorists and the appeal of 3D-printed weapons*. The Hague: International Centre for Counter-Terrorism.

Vergani, M., Iqbal, M., Ilbahar, E., & Barton, G. (2018). The three Ps of radicalization: Push, pull and personal. A systematic scoping review of the scientific evidence about radicalization into violent extremism. *Studies in Conflict & Terrorism, 43*(10), 854–886.

Vidino, L., & Brandon, J. (2012). *Countering radicalization in Europe*. London: International Centre for the Study of Radicalisation and Political Violence (ICSR).

Viki, G., Winchester, L., Titshall, L., & Chisango, T. (2006). Beyond secondary emotions: The infra-humanisation of groups using human-related and animal-related words. *Social Cognition, 24*, 753–775.

Wallner, C. (2021). *Against the Clock: Can the EU's new strategy for terrorist content removal work?* London: Royal United Services Institute.

Ware, J. (2019). *Siege: The Atomwaffen Division and rising far-right terrorism in the United States*. The Hague: International Centre for Counter-Terrorism.

Ware, J. (2020). *Testament to murder: The violent far-right's increasing use of terrorist manifestos*. The Hague: International Centre for Counter-Terrorism.

Ware, J. (2021). *Remembering Utoya, 10 years later: Enduring lessons for counter-terrorism*. The Hague: International Centre for Counter-Terrorism.

Warner, B. R. (2010). Segmenting the electorate: The effects of exposure to political extremism online. *Communication Studies, 61*(4), 430–444.

Watson, A., & Lupton, D. (2022). Remote fieldwork in homes during the COVID-19 pandemic: Video-call ethnography and map drawing methods. *International Journal of Qualitative Methods Volume, 21*, 1–12.

West, L., Juneau, T., & Amarasingam, A. (2021). *Stress tested: The COVID-19 pandemic and Canadian national security.* Calgary: University of Calgary Press.

Whitehouse, H. (2016). What motivates extreme self-sacrifice? *Pacific Standard, 9*(2), 26–31.

Whitehouse, H. (2018). Dying for the group: Towards a general theory of extreme self-sacrifice. *Behavioral and Brain Sciences, 41*, 1–62.

Whitehouse, H. (2021). *The ritual animal: Imitation and cohesion in the evolution of social complexity.* Oxford: Oxford University Press.

Whitehouse, H., & Lanman, J. A. (2014). The ties that bind us: Ritual, Fusion, and Identification. *Current Anthropology, 55*(6), 674–695.

Whitehouse, H., Jong, J., Buhrmester, M. D., Gómez, Á., Bastian, B., Kavanagh, C. M., Newson, M., Matthews, M., Lanman, J. A., McKay, R. & Gavriletset, S. (2017). The evolution of extreme cooperation via shared dysphoric experiences. *Scientific Reports, 7*(44292), 1–10.

Whitehouse, H., McQuinn, B., Buhrmester, M., & Swann, W. (2014). Brothers in arms: Libyan revolutionaries bond like family. *Proceedings of the National Academy of Sciences USA, 111*, 17783–17785.

Wiktorowicz, Q. (2005). *Radical Islam rising: Muslim extremism in the West.* London: Rowman & Littlefield Publishers.

Wildon, J., & Argentino, M.-A. (2021). *QAnon is not dead: New research into Telegram shows the movement is alive and well*. London: GNET Insights.

Wilson, S. M., & Peterson, L. C. (2002). The anthropology of online communities. *Annual Review of Anthropology, 31*, 449–467.
Winter, J. (2019, 1 August). Exclusive: FBI document warns conspiracy theories are a new domestic terrorism threat. *Yahoo News*.
Wong, J. C. (2020, 11 August). Revealed: QAnon Facebook groups are growing at a rapid pace around the world. *The Guardian*.
Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. Third IEEE International Conference on Data Mining, Melbourne.
YouTube (2023). *Hate speech policy.* YouTube Help Center. https://support-google.com/youtube (retrieved 10 June 2023).
Yuki, M. & Schug, J. (2012). Relational mobility: A socioecological approach to personal relationships. In O. Gillath, G. Adams, & A. Kunkel (Eds.) *Relationship Science: Integrating Evolutionary, Neuroscience, and Sociocultural Approaches* (pp. 137-151). American Psychological Association.
Zgryziewicz, R. (2018). *Violent extremism and communications.* Riga: NATO Strategic Communications Centre of Excellence.
Zihiri, S., Lima, G., Han, J., Cha, M., & Lee, W. (2022). QAnon shifts into the mainstream, remains a far-right ally. *Heliyon, 8*(2), e08764.
Zuckerman, E. (2019). QAnon and the emergence of the Unreal. *MIT Journal of Design and Science, 6*. https://doi.org/10.21428/7808da6b.6b8a82b9.

List of Figures and Tables

Figure 1: Proposed Relationships Between Group Alignment and Behavioural Outcome 21
Figure 2: Proposed Relationship Between Fusion and Violence 23
Figure 4: Group Alignment Detected in Analysed Manifestos 114
Figure 5: Other Variables of Interest Detected in the Analysed Manifestos............................118

Table 1: Overview of Scraped Online Groups 50
Table 2: Manifesto Coding 103
Table 3: Narrative Classifications and Definitions 104
Table 4: Intercoder Reliability (ICR) Analysis Results: Experts and Non-Experts......................110
Table 5: Detailed Overview of Narratives Detected in Manifestos 128
Table 6: Linguistic Proxies Detected in Manifestos 129
Table 7: Detected Linguistic Markers Across Manifestos (in %)...134
Table 8: Statistical Relevance of Linguistic Markers 136
Table 9. Overview of Selected Online Groups 153
Table 10. Weight Criteria for Violence Risk Index 155
Table 11. Quantitative Analysis Results 157
Table 12: QAnon Group Selection..182
Table 13: Quantitative Analysis Results (in %)...184
Table 14: Qualitative Analysis Results 186

Appendices

Appendix 1: R Codes for Manifesto Analysis

```
#FUSION
fusion <- list ("Brother","sister","family","motherland","our
blood","fatherland","sons","daughters","kin","my people","my race","our
people","European race","ancestry","ancestor","descendant","fellow", “brethren”,
“comrades”)

#VIOLENCE
violence <-
list("kill","hang","bomb","shoot","slaughter","exectued","execution","punish","death
penalty","massacre","destroy","must attack","must
fight","revenge","retribution","eradicate","starve","die","torture","behead","burn","br
ing death to","give them
hell","weapon","firearm","assassinate","gun","rifle","knife","grenade","brutal
steps","molotov","jihaad","jihad","set fire","revolution","forcible
overthrow","flamethrowers","M1-16","ammonium nitrate")

#IDENTIFICATION
identification1 <-
list("\\bwe\\b","\\bus\\b","\\bour\\b","\\bthey\\b","\\bthem\\b","\\btheir\\b")
identification2 <- list("\\bI\\b","\\bme\\b","\\bmy\\b","\\byou\\b","\\byour\\b")

#SLURS
slurs <- list("kike","nigger","negro","dirty jew","spic","fag","goyim","golem","the
jew","global jewry", “pajeet”, “bitch”, “whore”)

#DEMONISATION
demonisation <-
list("traitor","evil","enemy","corrupt","vicious","barbaric","depraved","vile","puppets"
,"perversion","blood libel","pervert","pedo","blood
libel","crime","cruel","bloody","genocidal","sinful","deceitful","invader","poison","par
asite","menace","brutal","ruthless","bloodsucking","dirty","deceptive","treacherous","
poisonous","oppressive","oppressor","shird","unbeliever","immoral","jahili","pollute",
"demolish","shake the foundations","dar ul-
harb","arrogant","mischievous","criminal","deceivers","liars")
```

```
#DEHUMANISATION
dehumanisation <- list("animal","plague","impure","brute","dog","lower iq","lower
being","inferior","squalid","parasitic","parasite","creature","trash","filth","vermin","sp
ider","devil","monster","beast","reptile","reptiloid","femoid","reptilian","snake","cock
roach","beneath human skin","sub human","anti-
human","disease","savage","infest","breed","locust","monkey","gorilla","rat","microb
e","satan","cancer", “scum”)

#EXISTENTIAL THREAT
existentialthreat <- list("subjected
to","coerced","brainwashed","exterminated","brutalised","raped","terrorised","ravag
ed","extinction","replacement","genocide","robbed","subjugate","make war upon my
people","destroy","subvert","overwhelmed","under siege","demographic
siege","disenfranchise","assault","kill us","kill our","kill my","running out of time","run
out of time","last chance","enslavement","enslaved","suffer","plunder","condemned
to death","destruction of all mankind","at the brink
of","endanger","annihiliation","decay")

#BELIEF IN CONSPIRACY
conspiracy <-
list("betray","betrayal","sell","sold","collude","conspire","fake","fraud","corruption","
corrupt","zog","great replacement","white genocide","kalergi","pedo
elites","NWO","illuminati","inside job","Eurabia")

#BELIEF IN INEVITABLE WAR
inevitablewar1 <- list("war","battle","fight","jihad","jihaad","collapse","conflict")
inevitablewar2 <- list("imminent","inevitable","looming","start", “begin”, "already",
“heading for”, “ongoing”, “stage”, “phase”, “when”, “has been”, “likely”, “predict”,
“expect”, “will happen”, “has begun”, “current”, “impending”)

#VIOLENCE JUSTIFICATION
violencejustification<- list("pre-emptive","defend","protect","self-defense","self-
defence","forced to fight","no longer ignore","act of
defense","purified","purify","need for war","need for violence","need for
jihaad","need for jihad","reasons for jihaad","reasons for war","reasons for
jihad","reasons for fight","reasons for violence","struggle is imposed","natural
struggle","cannot co-exist")

#MARTYRDOM NARRATIVE
martyr <- list("die in glory","sacrifice","knight","martyr","die selflessly","dying
selflessly","protecting our people","immortal","preserve","act of
```

```
preservation","defend the world of the Lord","defending the work of the Lord","stand
guard","standing guard","the herald","release mankind from","free from","freed
from")

#VIOLENT ROLE MODEL
violentrolemodel1 <-
list("breivik","tarrant","hitler","crusius","rodger","baillet","earnest","minassian","mcve
igh", “christchurch”, “poway” “el paso”)
violentrolemodel2 <- list("hero","role
model","saint","inspire","inspiration","inspiring","support","influenced")

#HOPELESSNESS OF ALTERNATIVE SOLUTIONS
hopelessness1 <-
list("democracy","democratic","peacful","political","system","politics","dialogue","pas
sivity")
hopelessness2 <- list("meaningless","weak","fail","end","vanish","man-
made","flawed","jahili","given up")

#CATEGORIES LIST
categorieslist<-
c("fusion","violence","slurs","demonisation","dehumanisation","existentialthreat","co
nspiracy","violencejustification","martyr", "violentrolemodel_", "identification%",
"hopelessness_", "inevitablewar")


#MATRIX
library(quanteda)
library(stringr)
library(parallel)

manifestomatrix<- matrix(,nrow=length(categorieslist)+2,ncol = length(manifestofiles))
manifestomatrix<- matrix(,nrow=length(categorieslist)+2,ncol = length(manifestofiles))
rownames(manifestomatrix)<- categorieslist
colnames(manifestomatrix)<-
lapply(manifestofiles,function(x){unlist(strsplit(unlist(strsplit(x, "\\."))[1], "\\/"))[2]})
for (i in 1:length(manifestofiles)) {
  manifesto=manifestofiles[i]
  name <- unlist(strsplit(unlist(strsplit(manifesto, "\\."))[1], "\\/"))[2]
  manifestosentences <- cbind(char_segment(get(name), pattern = ".", valuetype =
"fixed", pattern_position = "after"))
  for(j in 1:length(categorieslist)){
```

```
    category=categorieslist[j]
    if(grepl("_",category)){
      manifestofiltered =
manifestosentences[grep(paste(get(str_replace(category,"_","1")),collapse="|"),
manifestosentences, ignore.case=TRUE)]
      manifestofiltered =
manifestofiltered[grep(paste(get(str_replace(category,"_","2")),collapse="|"),
manifestofiltered, ignore.case=TRUE)]
      temp = length(manifestofiltered)/length(manifestosentences)*100
      print(length(manifestofiltered))
      manifestomatrix[j,i]<-ifelse(is.null(temp),0,temp)
       }else if(grepl("%",category)){
      manifestofiltered =
manifestosentences[grep(paste(get(str_replace(category,"%","1")),collapse="|"),
manifestosentences, ignore.case=TRUE)]
      manifestofiltered2 =
manifestosentences[grep(paste(get(str_replace(category,"%","2")),collapse="|"),
manifestosentences, ignore.case=TRUE)]
      temp<-length(manifestofiltered)/length(manifestofiltered2)
      manifestomatrix[j,i]<-ifelse(is.null(temp),0,temp)
    }else{
      manifestofiltered = manifestosentences[grep(paste(get(category),collapse="|"),
manifestosentences, ignore.case=TRUE),]
      temp = length(manifestofiltered)/length(manifestosentences)*100
      print(length(manifestofiltered))
      manifestomatrix[j,i]<-ifelse(is.null(temp),0,temp)
    }
    write.csv(manifestofiltered,paste0("Filtered Manifestos/",name,"_",category,".csv"))
  }
  manifestomatrix[j+1,i]<-length(manifestosentences)
  manifestomatrix[j+2,i]<-mean(mclapply(manifestosentences,ntoken))
}
write.csv(manifestomatrix,"CompleteManifestoMatrix.csv")
```

Appendix 2: R Codes for Online Group Analysis

```
#FUSION
data_filtered_fusion = QAnon[grep("Brother|sister|family|motherland|our
blood|fatherland|sons|daughters|kin|my people|my race|our people|European
race|ancestry|ancestor|descendant|fellow", QAnon$message, ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_fusion)/nrow(QAnon)

#VIOLENCE
data_filtered_violence =
QAnon[grep("kill|hang|bomb|shoot|slaughter|exectued|execution|punish|death
penalty|massacre|destroy|must attack|must
fight|revenge|retribution|eradicate|starve|die|torture|behead|burn|bring death to|
give them hell|weapon|firearm|assassinate|gun|rifle|knife|grenade|brutal
steps|molotov|jihaad|jihad|set fire|revolution|forcible
overthrow|flamethrowers|M1-16|ammonium nitrate", QAnon$message,
ignore.case=TRUE, useBytes = TRUE),]
nrow(QAnon)
nrow(data_filtered_violence)/nrow(QAnon)

#SLURS
data_filtered_slurs = QAnon[grep("kike|nigger|negro|dirty
jew|spic|fag|goyim|golem|the jew|global jewry", QAnon$message,
ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_slurs)/nrow(QAnon)

#DEMONISATION
data_filtered_demonisation = QAnon
[grep("traitor|evil|enemy|corrupt|vicious|barbaric|depraved|vile|puppets|perversio
n|blood libel|pervert|pedo|blood
libel|crime|cruel|bloody|genocidal|sinful|deceitful|invader|poison|parasite|menace
|brutal|
ruthless|bloodsucking|dirty|deceptive|treacherous|poisonous|oppressive|oppressor
|shird|
unbeliever|immoral|jahili|pollute|demolish|shake the foundations|dar ul-
harb|arrogant|mischievous|criminal|deceivers|liars", QAnon$message,
ignore.case=TRUE),]
nrow(QAnon)
```

```
nrow(data_filtered_demonisation)/nrow(QAnon)

#DEHUMANISATION
data_filtered_dehumanisation =
QAnon[grep("animal|plague|impure|brute|dog|lower iq|lower
being|inferior|squalid|parasitic|parasite|creature|trash|filth|vermin|spider|devil|m
onster|beast|reptile|reptiloid|femoid|reptilian|snake|cockroach|beneath human
skin|sub human| anti-
human|disease|savage|infest|breed|locust|monkey|gorilla|rat|microbe|satan|canc
er", QAnon$message, ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_dehumanisation)/nrow(QAnon)

#EXISTENTIAL THREAT
data_filtered_existentialthreat = QAnon[grep("subjected
to|coerced|brainwashed|exterminated|brutalised|raped|terrorised|ravaged|extincti
on|replacement|genocide|robbed|subjugate|make war upon my
people|destroy|subvert|overwhelmed|under siege|demographic
siege|disenfranchise|assault|kill us|kill our|kill my|running out of time|run out of
time|last chance|enslavement|enslaved|suffer|plunder|condemned to
death|destruction of all mankind|at the brink of|endanger|annihiliation|decay",
QAnon$message, ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_existentialthreat)/nrow(QAnon)

#BELIEF IN CONSPIRACY
data_filtered_conspiracybelief =
QAnon[grep("betray|betrayal|sell|sold|collude|conspire|fake|fraud|corruption|corr
upt|zog|great replacement|white genocide|kalergi", QAnon$message,
ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_conspiracybelief)/nrow(QAnon)

#BELIEF IN INEVITABLE WAR
data_filtered_inevitablewar =
QAnon[grep("war|battle|fight|jihad|jihad|collapse|conflict", QAnon$message,
ignore.case=TRUE),]
data_filtered_inevitablewar = data_filtered_inevitablewar
[grep("imminent|inevitable|looming|started|already", QAnon$message,
ignore.case=TRUE),]
nrow(QAnon)
```

```
# VIOLENCE JUSTIFICATION
data_filtered_violencejustification =
QAnon[grep(paste(violencejustification,collapse="|"), QAnon$message,
ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_violencejustification)/nrow(QAnon)

#MARTYRDOM NARRATIVE
martyr_dist <- function (message) {
  return(min(unlist(lapply(martyr, function(pattern) {
    drop(adist(pattern, message, partial = TRUE)) / nchar(pattern)})))))}
QAnon$martyrdom_distance = unlist(lapply(QAnon$message, martyr_dist))

data_filtered_martyrdom = QAnon[QAnon$martyrdom_distance<0.15,]
nrow(QAnon)
nrow(data_filtered_martyrdom)/nrow(QAnon)

#VIOLENT ROLE MODEL
data_filtered_violentrolemodel =
QAnon[grep("breivik|tarrant|hitler|crusius|rodger|baillet|earnest|minassian|mcveig
h", QAnon$message, ignore.case=TRUE),]
data_filtered_violentrolemodel = data_filtered_violentrolemodel[grep ("hero|role
model|saint|inspire|inspiration|inspiring|support|influenced", QAnon$message,
ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_violentrolemodel)/nrow(QAnon)

#HOPELESSNESS OF ALTERNATIVE SOLUTIONS
data_filtered_hopelessness = QAnon
[agrep("democracy|democratic|peacful|political|system|politics|dialogue|passivity",
QAnon$message, ignore.case=TRUE),]
data_filtered_hopelessness = data_filtered_hopelessness [grep
("meaningless|weak|fail|end|vanish|man-made|flawed|jahili|given up",
QAnon$message, ignore.case=TRUE),]
nrow(QAnon)
nrow(data_filtered_holelessness)/nrow(QAnon)
```

Appendix 3: Detailed ICR Analysis Results

Figure 1: Expert Intercoder Reliability Analysis: Initial vs. Revised Results

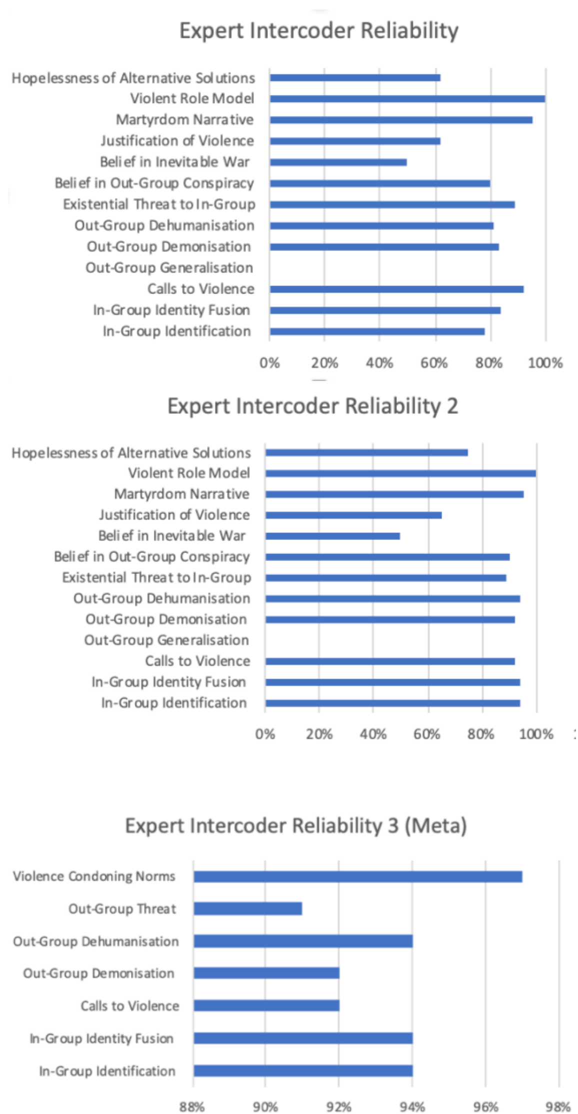
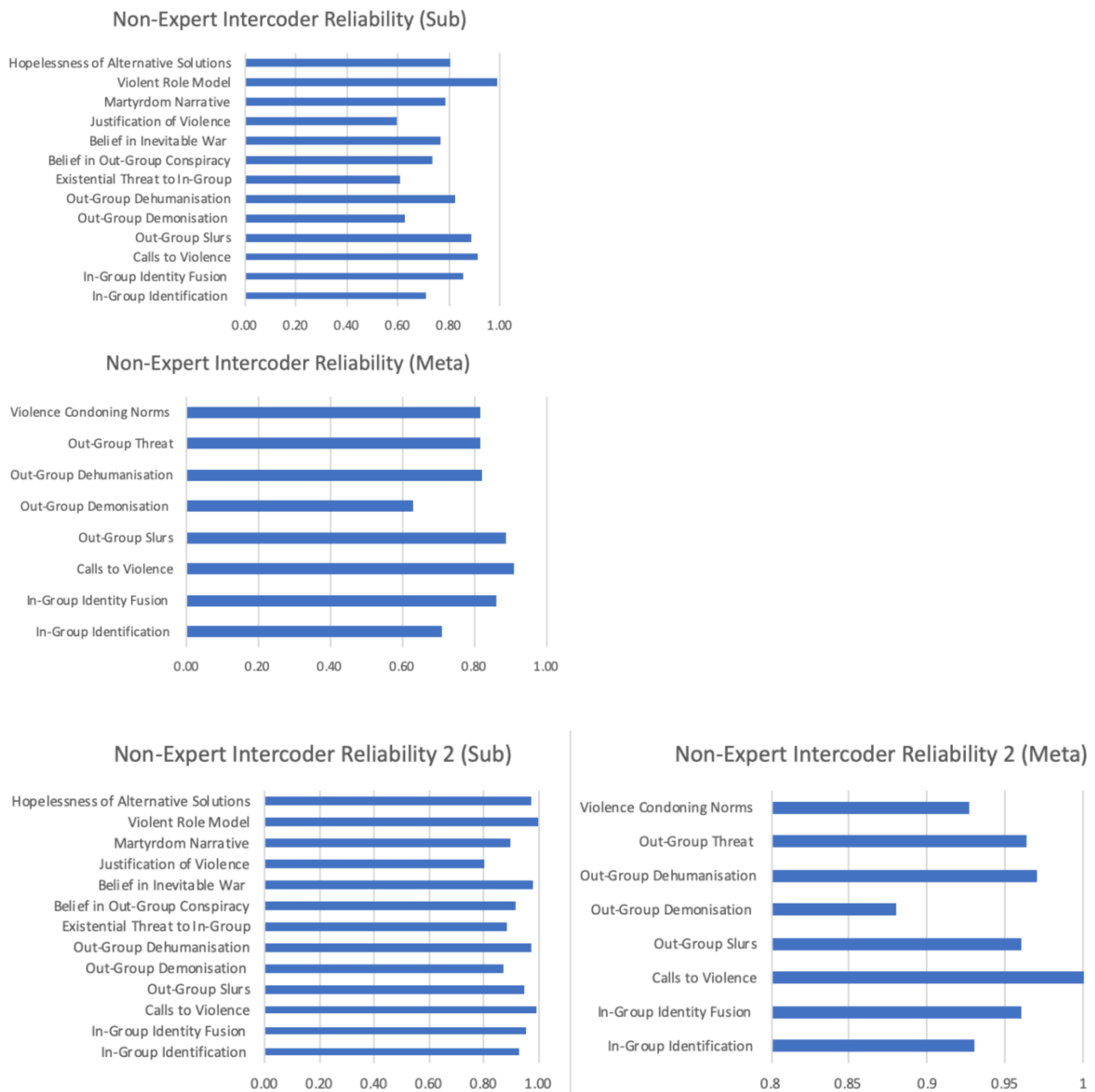


Figure 2: Non-Expert Intercoder Reliability Analysis: Initial vs. Revised Results



Category	***Mean Expert ICR***	***Mean Non-Expert ICR***
In-Group Identification	94%	93%
In-Group Identity Fusion	94%	96%
Calls to Violence	92%	100%
Out-Group Slurs	N/A	96%
Out-Group Demonisation	92%	88%
Out-Group Dehumanisation	94%	97%
Out-Group Threat	91%	96%
Violence Condoning Norms	97%	93%

Appendix 4: Critical Values of the Mann-Whitney U (two-tailed testing)

n_2	α	n_1																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	3	3
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	.05	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	.01	--	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	.05	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
	.01	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	.05	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	.01	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	.05	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	.01	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	.05	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	.01	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	.05	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	.01	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	.05	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	.01	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	.05	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	.01	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	.05	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	.01	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	.05	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	.01	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	.05	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	.01	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	.05	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	.01	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	.05	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
	.01	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105