



## Measuring socio-psychological drivers of extreme violence in online terrorist manifestos: an alternative linguistic risk assessment model

Julia Ebner, Christopher Kavanagh & Harvey Whitehouse

**To cite this article:** Julia Ebner, Christopher Kavanagh & Harvey Whitehouse (2024) Measuring socio-psychological drivers of extreme violence in online terrorist manifestos: an alternative linguistic risk assessment model, Journal of Policing, Intelligence and Counter Terrorism, 19:2, 125-143, DOI: [10.1080/18335330.2023.2246982](https://doi.org/10.1080/18335330.2023.2246982)

**To link to this article:** <https://doi.org/10.1080/18335330.2023.2246982>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 23 Aug 2023.



[Submit your article to this journal](#)



Article views: 2987



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

# Measuring socio-psychological drivers of extreme violence in online terrorist manifestos: an alternative linguistic risk assessment model

Julia Ebner , Christopher Kavanagh  and Harvey Whitehouse 

Department of Anthropology and Museum Ethnography, University of Oxford, Oxford, UK

## ABSTRACT

This paper develops a novel method of assessing the risk that online users will engage in acts of violent extremism based on linguistic markers detectable in terrorist manifestos. A comparative NLP analysis was carried out across fifteen manifestos on a scale from violent terrorist to non-violent politically moderate. We used a dictionary approach to measure the statistical significance of narratives previously identified in terrorism literature in predicting violence. The NLP analysis confirmed our research hypothesis, finding that the linguistic markers of identity fusion (an extreme form of group alignment whereby personal and group identities become functionally equivalent), dehumanising language towards the out-group and violence condoning norms were statistically significantly higher in manifestos of authors who engaged in acts of violent extremism. Building on our prior qualitative text analysis of terrorist manifestos, this study is among the first to offer a statistical analysis of the narrative patterns and associated linguistic markers distilled from terrorist manifestos. Beyond its academic contribution, the assessment framework presented here might assist security and counter-terrorism professionals in using psycholinguistic indicators to estimate the risk that online users will engage in offline violence and to make decisions on internal resource allocation in ongoing investigations.

## ARTICLE HISTORY

Received 18 January 2023


Accepted 30 July 2023

## KEYWORDS

Terrorist manifestos; violent extremism; violence risk assessment; linguistic predictors; identity fusion

## Introduction

Acts of violent extremism are often preceded by manifestos published online (Ware, 2020). But not all manifestos are followed by violence. Can we predict on the basis of the language used in manifestos which ones are likely to be followed by acts of murderous self-sacrifice and which ones can be safely ignored? The high number of violent threats made online has exacerbated the needle-in-a-haystack problem faced by counter-terrorism officials when making decisions about resource allocation and risk

**CONTACT** Julia Ebner  [Julia.ebner@sjc.ox.ac.uk](mailto:Julia.ebner@sjc.ox.ac.uk)  51/53 Banbury Road, Oxford, OX2 6PE, UK  @Julie\_renbe  
 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/18335330.2023.2246982>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

management. Not everyone who makes explicit threats to violence in online forums will translate their words into action. Likewise, not everyone who will commit an act of extreme violence threatens to do so beforehand. Our hypothesis is that perpetrators of violence often leave psychological traces in their online communications that inadvertently give away their violent intentions. This study seeks to contribute to a growing body of literature that attempts to test whether the occurrence of narrative patterns and linguistic markers can predict subsequent acts of extreme violence (Cohen, Johansson, Kaati, & Mork, 2014; Kaati, Shrestha, & Cohen, 2016; Van der Vegt, Kleinberg, & Gill, 2023). We argue that combining an evidence-based framework rooted in social psychology with computational linguistics can complement and enhance existing risk assessment frameworks used in the fields of counter-terrorism and predictive policing.

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 (Federal Bureau of Investigation, 2022; HM Government, 2015; The Council of Europe, 2022). 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, Comerford, Guhl, Baldet, & Colliver, 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 live action role plays (LARPS) and credible threats (Ebner, 2020; Radicalisation Awareness Network, 2021).

The aim of this study is to explore an alternative framework and its potential applications in detecting and monitoring groups and individuals at risk of committing extreme acts of pro-group violence. Our analysis presented here is theoretically grounded in literature on identity fusion (hereafter 'fusion' for short), an extreme form of group alignment characterised by a porous boundary between personal and group identities (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, McQuinn, Buhrmeister, & Swann, 2014), Indonesian religious fundamentalists (Kavanagh, Kapitány, Putra, & Whitehouse, 2020), Cameroonian herders and farmers (Buhrmeister, Zeitlyn, & Whitehouse, 2020), as well as British and Brazilian football hooligans (Newson, 2019). This study seeks to demonstrate the potential for psychologically grounded early warning systems that use a fusion-based linguistic model to prevent extreme violence. Compared to other forms of group cohesion such as in-group identification, fusion is a more robust predictor of violence. While strong group identification often co-occurs with high levels of identity fusion, previous research suggests that in the presence of threat fusion rather than identification is more often a potent driver of violent extremism (Swann & Buhrmeister, 2015; Whitehouse, 2018) In social psychology in-group identification describes an individual's sense of belonging to a defined group (Pennebaker & Chung, 2008).

An important advantage of our framework is that the examined key variables are not influenced by ideological or cultural factors. The framework can therefore be applied in different demographic, economic, cultural, and religious contexts. As the relevant variables tend to be revealed inadvertently 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. We do not propose to replace existing risk assessment models used by the intelligence community but to add an analytical layer that is grounded in relevant psychological theory.

This paper uses the term ‘ideological extremism’ drawing on J. M. Berger (2018)’s understanding of extremism as ‘the belief that an in-group’s success or survival is inseparable from the need for hostile action against an out-group’. In line with J.M. Berger’s framing, we apply the Institute for Strategic Dialogue (ISD)’s definition of extremism as ‘the advocacy of a system of belief that claims the superiority and dominance of one identity-based “in-group” over all “out-groups”’, and ‘advances a dehumanising “othering” mind-set incompatible with pluralism and universal human rights’ (ISD, 2022). The category ‘violent extremism’ will be used in this paper to refer to self-sacrificial pro-group violence. This may include terrorism, a term that we will use to describe acts of violent extremism that have been legally classified as terrorism.

Over a decade ago, Monahan (2012) concluded that evidence of reliable individual risk factors for terrorism remained scarce. Since then, many studies have attempted to make progress on the question what distinguishes violent from non-violent extremists (See for example Becker, 2019; Knight, Woodward, & Lancaster, 2017; Wolfowicz, Litmanovitz, Weisburd, & Hasisi, 2019 and Scrivens, 2022). Knight et al. (2017) compared a sample of 40 extremist case studies to find characteristics that distinguish violent extremists from non-violent extremists. Using a thematic analysis, they identified several factors that marked the cases of violent extremists, including underachievement (e.g. incongruence between academic education and employment status), turning point experiences (e.g. crisis) and environment with lack of boundaries for actions (‘open operating environment’). Becker (2019) analysed the relationship between social control, social learning and engagement in violent extremism, finding that weaker social control and stronger social learning of violence among radicalised individuals were more strongly associated with violent than with non-violent behaviour. Most recently, a study by Scrivens (2022) examined the posting patterns of violent and non-violent extremists on the prominent neo-Nazi forum Stormfront. His quantitative assessment showed that posts relating to extremist ideologies, personal grievances, and violent extremist mobilisation efforts were more prevalent in the posts of the non-extremist groups and control groups than in those of the violent groups. This suggests that online expressions of ideological extremism or indicators of violent mobilisation are not reliable factors in predicting actual violence.

A growing number of researchers have focused their attention on studying lone actor terrorists to detect patterns in the profiles, language and networks of the involved individuals. Meloy, Roshdi, Glaz-Ocik, and Hoffmann (2015) developed the so-called Terrorism Radicalization Assessment Protocol (TRAP-18), an investigative template that outlines eighteen factors to determine the risk of individual terrorism. Gill, Horgan, and Deckert (2014) found that no uniform profile exists based on their assessment of the sociodemographic networks, antecedent behaviours and motivations (including ideologies and

grievances) of 119 lone-actor terrorists. Their study did suggest, however, that lone-actor terrorism was often preceded by detectable activities within the wider social movements or groups the terrorists associated with. Studies by Horgan, Gill, Bouhana, Silver, and Corner (2016) and Clemmow, Gill, Bouhana, Silver, and Horgan (2020) raised the question whether the distinction between extremist and non-extremist lone actors makes sense at all. Their empirical research demonstrated the similarities between lone-actor terrorists and solo mass murderers and introduced a new overarching concept of lone-actor grievance-fuelled violence (LAGFV). This assessment is compatible with our fusion-based approach, which does not highlight ideology but rather focuses on identity as the key factor in determining proneness to violence.

Our paper is an attempt to help fill existing gaps in the literature, in particular relating to theoretically grounded linguistic analysis of terrorist manifestos and subsequent implications for the link between online discourse and offline violence. It combines qualitative and quantitative comparative analysis to design a new risk assessment framework that could be of use to security and counter-terrorism professionals. This study's results and practical implications should, however, be treated carefully. Sarma (2017) warned of the poor predictive value of individual violence risk assessments, and Van der Vegt et al. (2023) cautioned that the use of NLP in predictive policing is inherently marked by limitations due to data problems and low base rates. This study does not claim to provide a magic bullet for reliably singling out individual terrorists but it does seek to contribute to the development of theoretically grounded, evidence-based tools that can assist in the assessment of terrorism risk.

## Design

### *Data selection*

A total of fifteen manifestos were selected for the comparative NLP analysis on a scale from violent self-sacrificial to non-violent. The manifestos selected for statistical NLP analysis included the statements uploaded by high-profile terrorists prior to carrying out their attacks as well as prominent ideologically extreme and moderate political manifestos that were not followed by violence. In line with our terrorism definition outlined above, the manifesto selection categorised as 'violent self-sacrificial' included the statements published by the perpetrators of the 2011 Norway attacks, the 2014 Isla Vista killings (U.S.), the 2015 Charleston shooting (U.S.), the 2019 Christchurch mosque attacks (N.Z.), the 2019 attacks in Halle (Germany), Poway (U.S.), and El Paso (U.S.) as well as publications by NSDAP leader Adolf Hitler and the prominent jihadist revolutionary Sayyid Qutb. The non-violent selection covered the 'ideologically extreme' 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. The distinction between our non-violent categories 'ideologically extreme' and 'ideologically moderate' was applied based on the above-mentioned extremism definition.

The selected ideologically extreme (violent and non-violent) manifestos vary in their ideologies, covering anti-Muslim, anti-Semitic, misogynist and jihadist authors.

Meanwhile, the moderate (non-violent) manifestos include well-known feminist, environmental, and anti-racist texts. While our aim was to assess a wide range of manifestos on the ideological and tactical spectrum, all manifestos 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. All authors called for radical change, yet their choice of tactics to reach this goal differed. Due to the limited availability of terrorist manifestos, access restrictions to such documents, as well as researcher time considerations, we limited the number of analysed manifestos to fifteen.

### *Linguistic model*

The text-based framework applied in this study – including its choice of narrative categories and linguistic markers – was designed in a preceding study, which used comparative qualitative analysis to distil narrative and language patterns found in terrorist manifestos as opposed to manifestos that were not followed by violence (Ebner, Kavanaugh, & Whitehouse, 2022). Theoretically, our language analysis framework draws on the fusion-plus-threat model for predicting violent extremism (Whitehouse, 2018). Using this approach and theoretical framework, the fusion metrics as well as the other relevant variables, which will be outlined below, were tested in an intercoder reliability (ICR) assessment with two expert coders and twenty-four non-expert coders to confirm that they can be reliably identified and measured in written texts.

This study applies our newly developed linguistic framework in a systematic R-based Natural Language Processing (NLP) analysis, i.e. an automated computer-supported text analysis of our datasets. To assess the content of terrorist manifestos and compare it to non-violent (ideologically extreme and moderate) manifestos, we measure the degree of the following narrative categories in the R-based NLP code written for this study: identity fusion, existential threat to the in-group, out-group slurs, out-group demonisation, out-group dehumanisation, belief in conspiracy of the out-group, inevitable war narratives, justification of violence, martyrdom narratives, violent role models, and hopelessness of alternative solutions. Additionally, we tracked explicit calls to violence to understand how our framework variables interact with traditionally used language-based risk assessments. As mentioned above, the linguistic markers we used to measure the prevalence of the selected variables draw both on our qualitative research and existing literature on linguistic markers developed to measure forms of group cohesion and radicalisation drivers.

Identity fusion has been found to manifest itself in language in the form of images of kinship or metaphors of shared blood applied to other members of a group (Buhrmester, Fraser, Lanman, Whitehouse, & Swann, 2015; Swann et al., 2014). 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’ ‘fatherland’ might be used to talk about the in-group and/or fellow group members (Whitehouse & Lanman, 2014). Fully fused individuals might call members of their group ‘my white brothers’, ‘my native European sisters’, ‘brothers of Jihad’, ‘my brothers in blood’ or ‘my QAnon family’ (Ebner et al., 2022).

*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* 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, Markus, & Tait, 2018). It usually involves the framing of an out-group as bad, hostile, or threatening to the in-group. For example, studies explain 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 Dehumansation* 'involves viewing others as less than human', for example by describing them as or comparing them with animals (Vaes, Leyens, Paladino, & Miranda, 2012). Beyond the literal comparison with animals such as 'monkey', 'donkey', '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 demunisation (Viki, Winchester, Titshall, & Chisango, 2006).

*Existential Threat to In-Group* summarises the idea of the in-group being threatened with physical or symbolic collective annihilation (Hirschberger, Ein-Dor, Leidner, & Saguy, 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* denotes a functionally integrated mental system which 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, Clutton, & Dunn, 2019). *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, Conroy, & Ashwal, 2021).

*Justification of Violence* include rational or emotional reasonings of why resorting to violence is the best or only solution (Louis, Taylor, & Douglas, 2005; Spini, Elcheroth, & Fasel, 2008). For example, research highlighted group norms within jihadist groups that 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* 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). *Violent Role Models* may be mentioned in manifestos by invoking well-known perpetrators of genocidal violence as sources of inspiration (Cohen et al., 2014; Davey & Ebner, 2019). For example, authors might indicate support of previously successful terrorists by expressing identification, support or admiration (e.g. 'I admire', 'I salute', 'I support', naming someone 'Saint', 'God', etc.) for previous terrorists. *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 (Spears, 2010; Thomas & Louis, 2014).

*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). One of the challenges in research on expressed willingness to commit an act of violence lies in distinguishing between credible violence announcements and empty threats. A study that analysed the language of realised versus non-realised threats showed that adverbs of certainty and direct threats are more common in realised threats (stance markers of commitment, e.g. the word ‘will’), whereas value-laden words and paralinguistic items are more commonly found in non-realised threats (stance markers of emotion, e.g. the word ‘want’) (Gales, 2015). Nonetheless, many violent extremists never make explicit announcements or threats of violence before committing an atrocity. For this reason, this study seeks to provide insights into recurring socio-psychological patterns found in the language of would-be terrorists.

### **NLP analysis**

The NLP analysis of the manifestos was performed in R Studio – using the libraries *quanteda*, *stringr* and *parallel* (see Appendix 1 for the detailed R code) – 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 us to quantitatively examine the content in the different manifestos based on a dictionary approach, as well as the use of its integrated statistical functions. To maximise consistency in the coding categories, the intercoder reliability of the linguistic framework was previously tested with the help of two expert coders and twenty non-expert coders to the manual review. With an Intercode Reliability (ICR) score of over 90 percent for most narrative categories (Ebner et al., 2022), our framework achieved high reliability metrics.

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, Vishal, Suhane, Patro, & Mukherjee, 2020; Becker, Troschke, & Allington, 2021). For this reason, we chose to combine NLP-based analysis with a manual review of the filtered datasets to finetune our results. The filtered narrative datasets were exported from R and scanned manually for false positives. False positives were defined as sentences in the filtered narrative datasets that featured one or more linguistic markers but did not actually reflect our definitions of these narratives. For example, R could not distinguish

between fusion markers (kinship metaphors applied to the in-group) and mentions of real family members. Likewise, our computational results for dehumanising language directed towards the out-group also included mentions of real animals.

The total of over 100,000 sentences in 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 partially manual review:

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

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

Breivik's manifesto is a special case in point for 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, 2012). To reflect only words written by the author, a manual review of cited articles and copy-pasted passages was conducted. After articles by far-right ideologues such as Fjordman, Robert Spencer, and even journalists from media outlets, such as The Daily Mail, Telegraph, and the Weekly Standard, were deleted from the manifesto, a manual plagiarism check of Breivik's manifesto was conducted. This indicated that he had copied entire pages from books and online resources such as Wikipedia, often without citing them. After this adjustment Breivik's manifesto amounted to roughly 54,000 instead of 76,000 words (Table 1).

**Table 1.** Manifesto overview.

Author	Description	Total sentences	Coding category
Anders Behring Breivik	Manifesto of the Norway attacks, 2011	75,674 (adjusted: 53,526)	Violent self-sacrificial
Elliot Rodger	Manifesto of the Isla Vista killings, 2014	6,844	Violent self-sacrificial
Dylann Roof	Manifesto of the Charleston shooting, 2015	149	Violent self-sacrificial
Brenton Tarrant	Manifesto of the Christchurch mosque attacks, 2019	847	Violent self-sacrificial
Stephan Baillet	Manifesto of the Halle synagogue shooting, 2019	159	Violent self-sacrificial
John Earnest	Manifesto of the Poway synagogue attack, 2019	295	Violent self-sacrificial
Patrik Crusius	Manifesto of the El Paso attack, 2019	157	Violent self-sacrificial
Adolf Hitler	Mein Kampf, 1925	11,036	Violent self-sacrificial
Sayyid Qutb	Milestones, 1964	2,563	Violent self-sacrificial
Karl Marx and Friedrich Engels	Manifesto of the Communist Party, 1848	3,651	Ideologically extreme
Yusuf al-Qaradawi	The Lawful and Prohibited in Islam, 1960	8,197	Ideologically extreme
Fjordman	Defeating Eurabia, 2008	6,192	Ideologically extreme
Simone de Beauvoir	The Second Sex, 1949	13,052	Moderate
Martin Luther King Jr.	I Have a Dream, 1963	75	Moderate
Greta Thunberg	Our House is on Fire, 2019	63	Moderate

## Findings

What are the shared traits of terrorist writings, when compared to non-violent manifestos? The following visual matrix 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 2. shows that violent manifestos tended to be characterised by higher levels of identity fusion than non-violent manifestos of 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<sup>1</sup>:

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

or

$$U_1 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

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

We tested whether the difference was significant between:

- A) violent versus non-violent manifestos, and
- B) extreme versus non-extreme manifestos.

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 3. 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 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 2, the manifesto of the misogynist terrorist Elliot Rodger was an exception to the observed patterns. Based on our 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

**Table 2.** Detected linguistic markers across manifestos (in %).

[illegible]

**Table 3.** 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
n1	9	9	9	9	9	9	9	9	9	9	9	9
n2	6	6	6	6	6	6	6	6	6	6	6	6
R1	93.5	76.5	90	90	95.5	87	88	81	88.5	84	90	84
R2	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
U1	5.5	22.5	9	9	3.5	12	11	3	10.5	15	9	0
U2	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
n1	12	12	12	12	12	12	12	12	12	12	12	12
n2	3	3	3	3	3	3	3	3	3	3	3	3
R1	109.5	89	105	103	109.5	108	106.5	99	106.5	102	105	94.5
R2	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
U1	4.5	25	9	11	4.5	6	7.5	0	7.5	12	9	4.5
U2	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

(Involuntary Celibates) community he was loosely connected to (Hoffman, War, & Shapiro, 2020; Sugiura, 2021).

Our analysis presented here supports the underlying idea that one key invisible engine driving violent extremism is identity fusion when it occurs with other relevant variables, such as out-group dehumanisation and violence justification. Many individuals who espouse violence-confining norms (e.g. verbally justify violence or perceive alternative solutions as hopeless) never commit acts of violence. The same can be said of individuals who demonise or dehumanise perceived enemy groups. In isolation then each of the examined factors may prove harmless. However, when they co-occur, the components can react like chemical ingredients and become explosive.

However, it is important to stress that even when all the variables identified above as statistically significant are present there is no guarantee of violent actions. Too little is known about personal and collective violence barriers that would raise the threshold for an individual to actually commit an act of terrorism. Recent research into protective factors in violence risk assessment, points to a significant lack of empirical evidence on barriers to terrorism (Hewitt & Benjamin Lee, 2022). For instance, it is important for future research to explore and evaluate the impact of the human cost of self-sacrificial violence (e.g. individuals who have small children or are primary caregivers for family members) as an inhibiting factor. Some individuals might become more likely to commit an act of terrorism when the social, financial, and health-related costs are low (i.e. when they lack close family ties and care duties, face socio-economic instability or encounter severe health problems). Another potential inhibiting factor is empathy, although this may sometimes be strategically eroded by extremist groups by using dehumanising language towards the out-group (Čehajić, Brown, & González, 2009; Murrow & Murrow, 2015).

## Discussion

Our analysis suggests that there are detectable differences in the language used in the manifestos of violent versus non-violent extremists and that this is a more robust indicator than explicit declarations of violent intentions. Based on these findings, our study presents a new language assessment approach that could complement and enhance existing frameworks used by law enforcement agencies to prevent terrorism and other forms of extreme violence. Our theoretically grounded model therefore has the potential to offer an innovative approach in the field of countering violent extremism, which is confronted with rapidly changing tactics and radicalisation trends. By combining insights from cognitive anthropology, social psychology, and linguistics into an overarching linguistic framework, we want to contribute to a more holistic violence threat assessment framework that can help address urgent challenges in the counter-terrorism field. Our proposed NLP model 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 produced by the perpetrators, it seeks to make progress on the question of how to bridge the online-offline gap in research on violent extremism.[open-strick]

[close-strick]The framework is an attempt to address the challenges arising from the quickly evolving online extremism landscape. New user and content 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, hateful, abusive and violent words are frequently avoided by using code words or memes instead. (Calderón, Balani, & Taylor, 2021; NCTC, 2022) Moreover, as the lines between trolling and terrorism have become increasingly blurry, it has also become more difficult to distinguish between satire and credible threats. (Davey & Ebner, 2019; Ebner, 2020; Radicalisation Awareness Network, 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 – the FBI labels them 'salad bar ideologies' – means that ideologically rooted risk assessment frameworks are largely outdated (U.S. Government Publishing Office, 2022).

Examining recurring socio-psychological patterns in 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 on explicit threat expressions, known group membership, or specific ideological learnings. The new framework and findings presented here might furthermore complement existing academic studies, such as Gill et al. (2014), Meloy et al. (2015), Horgan et al. (2016), Clemmow et al. (2020) and Scrivens (2022). Our fusion-based approach could be applied to new datasets in combination with an analysis of sociodemographic networks, antecedent behaviours and motivations. For example, it would be to apply the framework to Gill et al.'s (2014) dataset of perpetrators of lone actor grievance-fuelled violence.

## Limitations

In a cautionary spirit, we wish to emphasise that this study is limited in its scope, as it only applied our model to a limited number of online manifestos uploaded prior to terrorist attacks, and compared them to non-violent political manifestos. The relatively small sample size and selection of manifestos mean that the implications of the findings presented here should not be overstated. While our datasets were large enough to allow for statistical analysis, the data problem highlighted in Van der Vegt et al. (2023) also applies to the study of terrorist manifestos, which is marked by both data scarcity and access restrictions. Our choice of non-violent manifestos included a range of prominent declarations that were all characterised by strong political or ideological views, an expressed urgency of their political or ideological message, and their targeting of public audiences. However, Martin Luther King Jr's *I have a Dream* and Greta Thunberg's *Our House is on Fire* were manifestos that were presented in the form of speeches rather than written publications. Another limitation lies in the large variation of word counts of the analysed texts. Future studies should expand the sample of comparative non-violent manifestos to include more examples of manifestos that meet the three above-mentioned criteria and provide a more powerful statistical analysis.

In this study we used R for the NLP analysis and coupled it with manual reviews. As expected, the NLP analysis provided a more nuanced perspective on the recurring narrative and linguistic patterns in terrorist manifestos than the previously conducted

qualitative manifesto analysis. As such, the R code used in this analysis is a useful starting point for further research. Although we expect the experiment presented here to be repeatable with other datasets, we acknowledge the need for the framework and associated R code to be further tested, developed, and updated in a constant feedback loop with the latest vocabulary, insider references, conspiracy myths and slurs used by extremist online communities. We recommend that follow-up projects test the framework in a diverse range of online groups, using ‘big data’ sources to further finetune the relevant components and linguistic markers. An idea for future research would be to include samples of chat logs left behind by convicted terrorists, which can then be compared and contrasted with the social media posts made by other members of the same online groups and chat forums who never resorted to violence.

One possible next step, would be to integrate our model with sophisticated machine learning language analysis tools such as CASM Technology’s Method52. The Java-based tool Method52, which was developed jointly by the Centre for the Analysis of Social Media (CASM) at the London-based think tank Demos, combines machine learning mechanisms with NLP capabilities and allows for the quantitative assessment of 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, 2016). 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, Ebner, Guhl, & Quent, 2018).

Finally, we recommend caution in the use of linguistic markers for fusion and related variables due to the high sensitivity of false positives and highly contextual nature of text-based expressions of socio-psychological drivers of violence. Manual reviews, from our perspective, 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 of 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 and al Qaradawi extensively covers Islamic laws for family life. Likewise, al Qaradawi explores the slaughtering of animals in relation to the teachings of Islam and De Beauvoir’s manifesto speaks in length about the biological differences between males and females in the animal world.

## Conclusion

Our analysis introduces a new language-based approach to determine the socio-psychological risk that an online user engages in extreme forms of pro-group violence. Our

analysis found that terrorist manifestos are marked by a statistically higher presence of linguistic indicators of identity fusion, as well as a higher occurrence of offensive, demonising and dehumanising language towards the out-group, and violence condoning norms (i.e. violence justification and hopelessness of alternative solutions). Based on previous studies and our statistical findings presented here, we argue that in isolation none of these factors may be indicative of proneness to violence. However, when linguistic markers for all components occur together, we anticipate a significantly higher risk of extreme violence.

This study's findings indicate the need for integrating socio-psychological markers, including identity-fusion, into existing violence risk assessment frameworks. Our fusion-based NLP framework can contribute to improving and expanding existing early warning systems used by law enforcement and security efforts. However, putting the research findings of this study into practice must proceed with caution to avoid undermining user privacy and criminalising socio-psychological phenomena. We recommend careful navigation of the ethical challenges associated with practical implementation of these research findings to avoid exacerbating already well-established ethical problems with many counter-terrorism interventions (Silke, 2019).

The importance of identity fusion in radicalisation pathways towards violence also suggest that identity-centred approaches should be prioritised in terrorism prevention and radicalisation intervention. One route may consist in 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 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 to 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 useful for intervention efforts aimed to encourage users to turn away from their fused identity.

Finally, we recommend that government units, intelligence agencies, police services, courts and tech firms make more data available to researchers. Access restrictions regarding the raw communication materials (e.g. chat logs) of convicted terrorists have posed a challenge for quantitative research into violence risk factors. To improve the predictive power of studies in the field of counter-terrorism, a closer collaboration between the intelligence community and academic researchers would highly be beneficial.

## Note

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

## Acknowledgements

The authors would like to express their gratitude to Philip Kreißel, data scientist at University of Bamberg, Hateaid and Volksverpetzer for his tech expertise, mentoring and advice in the NLP coding process.

## Data availability statement

Due to the sensitivity of the analyzed content, the researchers refrain 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.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the UK's Economic and Social Research Council and St John's College at Oxford University under the ESRC Grand Union Doctoral Training Partnership (DTP) Studentship and an Advanced Grant from the European Research Council (ERC) under the European Union's HORIZON EUROPE 2020 Research and Innovation Programme (#694986).

## ORCID

Julia Ebner  <http://orcid.org/0000-0002-1922-1204>

Christopher Kavanagh  <http://orcid.org/0000-0001-7545-8137>

Harvey Whitehouse  <http://orcid.org/0000-0001-6935-6724>

## Works Cited

- Alon, N., & Omer, H. (2005). *The psychology of demonization: Promoting acceptance and reducing conflict*. London: Routledge.
- al-Qaradawi, D. Y. (1960). *The lawful and prohibited in Islam*. Davao City: Wisdom International School for Higher Educational Studies (WISHES).
- Amanullah, Z., & Harrasy, A. (2017). *Between two extremes: Responding to Islamist and Tribalist messaging online in Kenya during the 2017 elections*. London: ISD.
- Anderson, L., & Lepore, E. (2013). Slurring words. *Noûs*, 47, 25–48. doi:10.1111/j.1468-0068.2010.00820.x
- Bale, J. (2007). Political paranoia v. political realism: On distinguishing between bogus conspiracy theories and genuine conspiratorial politics. *Patterns of Prejudice*, 41, 45–60. doi:10.1080/00313220601118751
- 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. *Association for Computational Linguistics*, 1028–1023.
- Becker, D. M., Troschke, H., & Allington, D. D. (2021). *Decoding antisemitism: An AI-driven study on hate speech & imagery online*. Berlin: TU Berlin.
- Becker, M. H. (2019). When extremists become violent: Examining the association between social control, social learning, and engagement in violent extremism. *Studies in Conflict & Terrorism* 44(12), 1104–1124.
- Berger, J. M. (2018). *Extremism. The MIT essential knowledge series*. Cambridge, MA: The MIT Press.
- 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. doi:10.3917/rfsp.615.0867
- Buhrmeister, M., Zeitlyn, D., & Whitehouse, H. (2020). Ritual, fusion and conflict. *Group Processes & Intergroup Relations*, 25(1), 298–311. doi:10.1177/1368430220959705

- Buhrmester, M., Fraser, W. T., Lanman, J. A., Whitehouse, J., & Swann, W. B. (2015). When terror hits home: Identity fused Americans who saw Boston bombing victims as "family" provided aid. *Self and Identity*, 14(3), 253–270. doi:10.1080/15298868.2014.992465
- Calderón, F. H., Balani, N., & Taylor, J. (2021). Linguistic patterns for code word resilient hate speech identification. *Sensors*, 21(23), 7859. doi:10.3390/s21237859
- CASM. (2016, May 26). *New Demos study reveals scale of social media misogyny*. Abgerufen am 3. December 2019 von Demos: <https://demos.co.uk/press-release/staggering-scale-of-social-media-misogyny-mapped-in-new-demos-study/>
- Čehajić, S., Brown, R., & González, R. (2009). What do I care? Perceived ingroup responsibility and dehumanization as predictors of empathy felt for the victim group. *Group Processes & Intergroup Relations*, 12, 715–729. doi:10.1177/1368430209347727
- Clemmow, C., Gill, P., Bouhana, N., Silver, J., & Horgan, J. (2020). Disaggregating lone-actor grievance-fueled violence: Comparing lone-actor terrorists and mass murderers. *Terrorism and Political Violence*, 34(3), 558–584. doi:10.1080/09546553.2020.1718661
- 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. doi:10.1080/09546553.2014.849948
- 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 (ISD).
- Davey, J., & Ebner, J. (2019). *The great replacement': The violent consequences of mainstreamed extremism*. London: Institute for Strategic Dialogue.
- de Beauvoir, S. (1949). *Le Deuxième Sexe*. Paris: Editions Gallimard.
- Ebner, J. (2020). *Going dark: The secret social lives of extremists*. London: Bloomsbury.
- Ebner, J., Kavanagh, C., & Whitehouse, H. (2022). Is there a language of terrorists: A comparative manifesto analysis. *Studies in Conflict and Terrorism*, 1–28.
- el-Husseini, R. (2008). Resistance, jihad, and martyrdom in contemporary Lebanese Shi'a discours. *Middle East Journal*, 62(3), 399–414. doi:10.3751/62.3.12
- Federal Bureau of Investigation. (2022, November 15). *What we investigate: FBI threat intimidation guide*. Von FBI. <https://www.fbi.gov/investigate/counterintelligence/threat-intimidation-guide#Eletronic-message%20threat> abgerufen
- Fielitz, M., Ebner, J., Guhl, J., & Quent, M. (2018). *Loving hate: Anti-Muslim extremism, radical Islamism and the spiral of polarization*. Jena, London, Berlin: IDZ and ISD.
- Fjordman. (2008). *Defeating Eurabia*. self-published book.
- Fraise, T. (2017). *Shades of jihad: Variation of military ethics between ISIS and alQaeda*. Paris: Sciences Po Kuwait Program Research Paper.
- Gales, T. (2015). Threatening stances: a corpus analysis of realized vs. non-realized threats. *Language and Law*, 2(2), 1–25.
- Gill, P., Horgan, J., & Deckert, P. (2014). Bombing alone: Tracing the motivations and antecedent behaviors of terrorists. *Journal of Forensic Sciences*, 59(2), 425–435. doi:10.1111/1556-4029.12312
- 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. doi:10.1080/09546550601054873
- Hewitt, J., & Benjamin Lee, S. M. (2022). *Protective factors in risk assessment: Practitioner perspectives*. London: Centre for Research and Evidence on Security Threats (CREST).
- Hirschberger, G., Ein-Dor, T., Leidner, B., & Saguy, T. (2016). How is existential threat related to inter-group conflict? Introducing the multidimensional existential threat (MET) model. *Front Psychology*, 7, 1877. doi:10.3389/fpsyg.2016.01877
- HM Government. (2015). *Counter-terrorism and security act 2015*. Von Legislation: <https://www.legislation.gov.uk/ukpga/2015/6/contents/enacted.abgerufen>
- Hoffman, B., War, J., & Shapiro, E. (2020). *Assessing the threat of incel violence*. St Andrew's: St Andrew's University.
- Horgan, J. G., Gill, P., Bouhana, N., Silver, J., & Corner, E. (2016). *Across the universe? A comparative analysis of violent behavior and radicalization across three offender types with implications for*

- criminal justice training and education*. Washington D.C.: Research report submitted to the U.S. Department of Justice.
- ISD. (2022). *Extremism definition*. Von institute for strategic dialogue. <https://www.isdglobal.org/extremism/abgerufen>
- Jeshion, R. (2013). Slurs and stereotypes. *Analytic Philosophy*, 54(3), 314–329. doi:10.1111/phib.12021
- Kaati, L., Shrestha, A., & Cohen, K. (2016). *Linguistic analysis of lone offender manifestos*. Von 2016 IEEE international conference on cybercrime and computer forensic (ICCCF). [https://www.foi.se/download/18.7fd35d7f166c56be0b1003b/1542623789571/Linguistic-analysis-of-lone\\_FOI-S-5650-SE.pdf](https://www.foi.se/download/18.7fd35d7f166c56be0b1003b/1542623789571/Linguistic-analysis-of-lone_FOI-S-5650-SE.pdf) abgerufen
- Kavanagh, C., Kapitány, R., Putra, I. E., & Whitehouse, H. (2020). Exploring the pathways between transformative group experiences and identity. *Frontiers in Psychology* 11, 1–16.
- King, M. L., Jr. (1963). I have a dream. *Speech*. Washington, DC.
- 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). doi:10.1371/journal.pone.0225098
- Knight, S., Woodward, K., & Lancaster, G. (2017). Violent versus non-violent actors: Empirical study of different types of extremism. *American Psychological Association* 4(4), 230–248.
- Kriner, M., Conroy, M., & Ashwal, Y. (2021). *Understanding accelerationist narratives: 'There is no political solution'*. London: GNET Report, Kings College.
- Kundnani, A. (2012). *The anti-Islamist: Anders Behring Breivik's manifesto*. The Hague: International Center for Counter-Terrorism (ICCT).
- Lifton, R. (1988). *The Nazi doctors: Medical killing and the psychology of genocide*. New York City: Basic Books.
- 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. doi:10.1177/1368430205056465
- Meloy, J. R., Roshdi, K., Glaz-Ocik, J., & Hoffmann, J. (2015). Investigating the individual terrorist in Europe. *Journal of Threat Assessment and Management*, 2(3-4), 140–152. doi:10.1037/tam0000036
- Miller, C. (2016). *From Brussels to Brexit: Islamophobia, xenophobia, racism and reports of hateful incidents on Twitter*. London: Demos.
- Miller-Idriss, C. (2021). *White supremacist extremism and the far right in the U.S.* Farmington Hills: Gale Primary Sources/Cengage.
- Monahan, J. (2012). The individual risk assessment of terrorism. *Psychology, Public Policy, and Law*, 18 (2), 167–205. doi:10.1037/a0025792
- Murrow, G. B., & Murrow, R. (2015). A hypothetical neurological association between dehumanization and human rights abuses. *Journal of Law and Biosciences*, 2(2), 336–364. doi:10.1093/jlb/lsv015
- NCTC. (2022). *Use of memes by violent extremists*. Washington, DC: NCTC, DHS, FBI.
- Newson, M. (2019). Football, fan violence, and identity fusion. *International Review for the Sociology of Sport*, 54(4). doi:10.1177/1012690217731293
- Pennebaker, J., & Chung, C. K. (2008). Computerized text analysis of Al-Qaeda transcripts. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 453–465). Newbury Park, CA: Sage.
- Radicalisation Awareness Network. (2021). *The gamification of violent extremism & lessons for P/CVE*. Luxembourg: Publications Office of the European Union.
- 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), 1–22.
- Sage Research Methods. (2016). *Digital social research: Centre for the analysis of social media (CASM)*. Abgerufen am 3. Dezember 2019 von <https://methods.sagepub.com/video/digital-social-research-centre-for-the-analysis-of-social-media-casm>
- Sarma, K. M. (2017). Risk assessment and the prevention of radicalization from nonviolence into terrorism. *American Psychologist*, 72(3), 278–288. doi:10.1037/amp0000121

- 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.
- Scrivens, R. (2022). Examining online indicators of extremism among violent and non-violent right-wing extremists. *Terrorism and Political Violence* 35(6), 1389–1409.
- Silke, A. (2019). *Routledge handbook of terrorism and counterterrorism*. Abingdon: Routledge.
- Spears, R. (2010). Group rationale, collective sense: Beyond intergroup bias. *British Journal of Social Psychology*, 49, 1–20. doi:10.1348/014466609X481308
- 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. doi:10.1111/j.1467-9221.2008.00673.x
- Sugiura, L. (2021). *The incel rebellion: The rise of the manosphere and the virtual war against women*. Bingley: Emerald Publishing Limited.
- Swann, W. B., & Buhrmeister, M. D. (2015). Identity fusion. *Current Directions in Psychological Science*, 24(1), 52–57. doi:10.1177/0963721414551363
- Swann, W. B., Buhrmester, M. D., Gómez, A., Jetten, J., Bastian, B., Vázquez, A., ... Finchilescu, G. (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. doi:10.1037/a0036089
- Technau, B. (2018). Going beyond hate speech: The pragmatics of ethnic slur terms. *Lodz Papers in Pragmatics*, 14(1), 25–43. doi:10.1515/lpp-2018-0002
- The Council of Europe. (2022, November 12). *Types of cyberviolence, cyberviolence at a glance*. Von CoE. <https://www.coe.int/en/web/cyberviolence/types-of-cyberviolence> abgerufen
- 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. doi:10.1177/0146167213510525
- Thunberg, G. (2019). Our house is on fire. *Forum insight*. Davos: World Economic Forum.
- U.S. Government Publishing Office. (2022). *The dynamic terrorism landscape and what It means for America*. Washington, DC: Hearing Before the Committee on Homeland Security, House of Representatives, 117th Congress.
- 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). doi:10.1080/10463283.2012.665250
- Van der Vegt, I., Kleinberg, B., & Gill, P. (2023). Proceed with caution: On the use of computational linguistics in threat assessment. *Journal of Policing, Intelligence and Counter Terrorism*, 18(2), 231–239. doi:10.1080/18335330.2023.2165137
- Veilleux-Lepage, Y. (2021). *CTRL, HATE, PRINT: Terrorists and the appeal of 3D-printed weapons*. The Hague: ICCT.
- 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. doi:10.1521/soco.2006.24.6.753
- Ware, J. (2020). *Testament to murder: The violent far-right's increasing use of terrorist manifestos*. The Hague: International Center for Counter Terrorism (ICCT).
- Whitehouse, H. (2018). Dying for the group: Towards a general theory of extreme self-sacrifice. *Behavioral and Brain Sciences*, 41, 1–62. doi:10.1017/S0140525X17000012
- Whitehouse, H., & Lanman, J. A. (2014). The ties that bind us: Ritual, fusion, and identification. *Current Anthropology*, 55(6), 674–695. doi:10.1086/678698
- Whitehouse, H., McQuinn, B., Buhrmeister, M., & Swann, W. (2014). Brothers in arms: Libyan revolutionaries bond like family. *Proceedings of the National Academy of Sciences*, 111(50), 17783–17785. doi:10.1073/pnas.1416284111
- Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2019). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalisation outcomes. *Journal of Quantitative Criminology* 3, 407–447.