

Extreme parallels: a corpus-driven analysis of ISIS and far-right discourse

Louisa Buckingham & Nusiebah Alali

To cite this article: Louisa Buckingham & Nusiebah Alali (2019): Extreme parallels: a corpus-driven analysis of ISIS and far-right discourse, Kōtuitui: New Zealand Journal of Social Sciences Online, DOI: [10.1080/1177083X.2019.1698623](https://doi.org/10.1080/1177083X.2019.1698623)

To link to this article: <https://doi.org/10.1080/1177083X.2019.1698623>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 11 Dec 2019.



[Submit your article to this journal](#)



Article views: 106



[View related articles](#)



[View Crossmark data](#)

Extreme parallels: a corpus-driven analysis of ISIS and far-right discourse

Louisa Buckingham * and Nusiebah Alali *

School of Cultures, Languages and Linguistics, University of Auckland, Auckland, New Zealand

ABSTRACT

In this study, we examine key psychological dimensions in the manifestos authored by the perpetrators of the Christchurch and Utøya massacres, the right-wing extremists Brenton Tarrant and Anders Breivik, and the ISIS propaganda magazine Rumiya. All texts were authored and disseminated virtually with the purpose of attracting or consolidating support, and justifying violent, discriminatory actions. While right-wing supremacist and extremist Islamist discourses are ostensibly ideologically opposed, previous research has posited the existence of ideational and emotive commonalities. We approach this from a corpus-linguistic perspective, and employ the software LIWC2015 and Wmatrix to explore the dominant psychological dimensions, semantic categories and keywords in these texts. We identify elements that contribute to the construction of a narrative of hate, peril and urgency, and discuss differences in the imagery used to construct these meanings and to appeal to different audiences. Whilst our analysis supports the existence of commonalities in ideological content and discursive strategies, our results identify differences in the target of hate in right-wing supremacist discourse and we differentiate between primarily Islamophobic and racist motives. Finally, we also discuss the limitations inherent in employing these software tools to analyse discourse in the Web 2.0 era.

ARTICLE HISTORY

Received 7 October 2019
Accepted 25 November 2019


KEYWORDS

Extremist discourse; right-wing supremacy; ISIS; corpus linguistics; psychological dimensions

Introduction

The massacre of 51 people at two mosques in Christchurch in March 2019 catapulted New Zealand into the orbit of transnational right-wing extremism. The lone-wolf profile of perpetrator Brent Tarrant, his apparent lack of affiliation to right-wing or extremist organisations and the form in which he broadcast and justified his actions, are reminiscent of the 2011 massacre in Norway. Perpetrator Anders Behring Breivik targeted two locations, a public office in Oslo and a holiday camp on the island of Utøya, in a consecutive manner, and released an electronic document (*2083: A European Declaration of Independence*), known later as his ‘manifesto’. Tarrant targeted individuals at two mosques

CONTACT Louisa Buckingham  l.buckingham@auckland.ac.nz

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/1177083X.2019.1698623>.

*Both authors contributed equally to this work.

© 2019 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-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

sequentially and disseminated his deeds and his 'manifesto' (*The Great Replacement*) through social media. His live-streaming broadcast of his actions enabled viewers to 'experience' the massacre in a manner that recalls online first-person shooter games, and illustrates how materials (both explanatory and experiential) can be disseminated and accessed in real-time, that is, synchronous to the violent act.

A central function of both manifestos was to provide a public explanation of the authors' actions, the logic behind the respective massacres and the choice of target for their hate: in Norway, politically engaged social democrats, and in New Zealand, practising Muslims. While Breivik targeted the social group that he considered responsible for the increased migration to Norway of individuals of non-European or non-Christian heritage, Tarrant intentionally targeted this religious community. Whilst the two perpetrators present themselves as 'authors', both manifestos contain material copied from third sources. The narratives are similar to those found in other forms of extreme right-wing discourse, and are thus 'embedded in ... the rhetoric of larger social movements' (Berntzen and Sandberg 2014, p. 760).

The Internet has enabled extremist groups an international outreach in terms of connectivity with likeminded, resource mobilisation and the dissemination of multimodal propaganda. While right-wing, European supremacist groups do not have a strong presence in Norway or New Zealand (see, for instance, Gilbert 2013), Internet de-territorialises an individual's political enculturation, and can upscale personal or local narratives of grievance to a transnational narrative of collective existential threat. Legal restrictions on the creation and dissemination of extremist material in one locality (e.g. the censoring of Tarrant's materials in New Zealand) can be evaded through online access to such materials disseminated abroad. The virtual diffusion of such propaganda (textual and audiovisual) has become a mechanism to cultivate support, foster identification and commitment, and encourage the transformation of followers into activists.

The high-choice media environment of the Internet, and particularly social media, can serve an enabling environment for radicalisation¹ as it can provide selective exposure to information (textual and visual), and reinforce existing beliefs and preferences (Hawdon et al. 2019). Empirical research has shown that the resultant potential echo-chamber effect is likely to be stronger among more radical or conservative sub-groups, and those uninterested in (mainstream) politics and who lack exposure to diverse views from family and acquaintances (O'Hara and Stevens 2015; Dubois and Blank 2018).

In their survey of US internet users, Costello and Hawdon (2018) noted a salience of right-wing extremist groups on the Internet and the importance of identification with an online hate-focused online group to the radicalisation process of individuals. Nevertheless, this behaviour is not specific to this target group. Investigations into the radicalisation of Islamist extremists have identified a similar process of disenchantment with one's current life, disengagement from one's immediate surrounds, and identification with multimodal (religious) supremacist narratives circulated in the Internet, which provide a sense of cultural security, self-identification, importance, purpose, and ultimately, empowerment (Cottee and Hayward 2011).

While it has not been possible to determine a causal relationship between exposure to extremist online content and radicalisation (Cottee and Cunliffe 2018), frequent exposure to hate online, close identification with (and interacting with) an online community of this persuasion, are associated with greater likelihood of the individual contributing to online

hate discourse (Hawdon et al. 2019). Further, atrocities depicted or described in extremist online materials have been enacted or replicated by ‘lone wolf’ individuals. The recent example involving the beheading of two Scandinavian tourists in Morocco in December 2018 resembled the widely publicised beheadings of captives by ISIS between 2014 and 2015. Alternatively, extremist material may contain instructions on undertaking terrorist acts. Issues 3 and 4 of the ISIS magazine *Rumiyah* provided detailed instructions on the use of a vehicle or a knife to attack pedestrians, and such acts have been undertaken in recent years by individuals who (retroactively) claimed to be ISIS sympathisers (e.g. the attack in Germany, December 2016, perpetrated by Anis Amri).

Previous studies have posited that the narrative propagated in religious and white supremacist extremist materials displays similarities, and indeed Holbrook’s (2013) close reading of extremist texts identified ideational and emotive commonalities. To investigate this further we propose a corpus-driven analysis of the conceptual meanings contained in extremist texts. We posit that such meanings can be identified through a cross-ideological comparative analysis of significant semantic and lexical features. To this end, we compare the manifestos authored by white supremacists Breivik and Tarrant (789, 908 and 17,160 words, respectively) and the 13 issues of the ISIS magazine *Rumiyah* (298,351 words),² and we (primarily) employ two different corpus analytic tools to facilitate methodological triangulation, as recommended in the discipline (e.g. Baker et al. 2008). Our study seeks to identify similarities and differences in the psychological-semantic content of these texts, and it explores how significant psychological-semantic constructs are expressed at a lexical level.

Extremist discourse

Recognised as a broad term entailing a variety of ideological characteristics, right-wing extremism has been described by Van der Valk (2013) as including three main strands: firstly, a resistance to the principle of equality of human beings and the espousal of the importance of belonging to a self-perceived ‘superior’ social group defined along the lines of race, ethnicity, culture, religion. The ensuing pronounced ingroup orientation, and negative outgroup orientation is manifested as actively pursued exclusionary behaviour towards outgroup individuals. The second strand entails resistance to the parliamentary democratic political system and processes, often combined with a claim to represent ‘the people’; and the final strand involves the preference for an authoritarian hierarchical organisation and strong undemocratic leadership.

Similarly, Liebman (1983) identified three principal elements of religious extremism. The first element is the desire to expand the scope of religious legislation, the development of the specifics of law, and rigidity in the interpretation of the law. The second element involves the call to socially isolate or exclude followers of other religions or co-religionists who do not conform to extremist ideology. The last element entails the rejection of cultural values and norms that are considered foreign to the sanctioned religious culture.

As the foregoing illustrates, extremist discourse, whether political or religious, relies on an exclusivity element. All religions (and arguably also political movements) entail an element of exclusivity, however, in their endeavour to position their interpretation as possessing ‘the essence and substance of true universal religion’ (Pratt 2015, p. 210). Pratt (2015) recognises three levels of exclusivity and acknowledges only the third as amounting

to extremism, in that it involves concerted resistance to others, including their elimination, relies on a narrative of superiority, and rejects any form of dialogic exchange.

The compulsion to explain actions, in particular, violence, to the general public has been found among politically motivated extremist individuals or organisations of very different persuasions. As elucidated by Lehr (2013), both Germany's Red Army Faction and Al-Qaida's leadership disseminated texts or video messages to explain or claim responsibility for terrorist acts. The resultant existence of material intended to provide insight into the logic behind the acts and, presumably, also gain sympathisers, enables an examination into the discursive strategies employed by extremists to rationalise their perspectives. Previous research has pointed to apparent discursive and ideological similarities in the texts produced by opposing extremist groups or individuals. Both forms of politically motivated extremists are ideologically aligned in their rejection of democracy and their resort to violence to attain their objectives (Van der Valk 2013). Holbrook (2013) points to ideational and strategic similarities; both groups perceive themselves as leading the battle against an existential threat to their values and identities. The strategy of emotional appeals to potential sympathisers is intended to invoke a sense of mutual grievances, shared cause, and duty to respond. A further similarity entails the importance of myths which have the function, according to Bieber (2002, p. 97) in his analysis of Serb nationalist discourse, of 'eliminating the historical separation between past and present by contemporizing the past or historicizing the present'. The interpretation of the past is thereby adjusted to accord with contemporary social and political narratives. In right-wing extremist discourse, frequent myths include the notion of racial or cultural homogeneity (and superiority).

In a similar vein, extremist Islamist discourse exploits master narratives, which Halverson, Corman, and Goodall (2011, p. 14) define as a 'transhistorical narrative that is deeply embedded in a particular culture', to create a well-crafted persuasive narrative. Halverson et al. (2011) identified around thirteen master narratives in Islamic culture that predominate in extremist Islamist discourse, and which include, inter alia, the Pharaoh, Jahiliya, the Hypocrites and the Crusaders. The success of these narratives in attracting the target audience stems from the fact that they are constructed within a grand narrative that Muslims use to make sense of the world and build connections between the past and present. Through systematic selective citations of Islamic sources i.e. the Qur'an, Hadith and Islamic history, extremist propaganda creates the illusion of a legitimate narrative. For instance, Mahood and Rane (2017) identified that the master narratives pertaining to Jahiliya and the Crusaders in ISIS propaganda are used to foment feelings of discrimination and victimisation among Muslims residing in the West and serve as a justification for terrorist acts against the culprit, 'the West'.

In her assessment of research on extremist discourse in the Internet, Conway (2017) identified the need for large-scale, systematic analyses. Analyses of extremist discourse to date have usually involved a close reading of a selected text (see e.g. Bertzen and Sandberg 2014), or a selective focus on particular words and phrases from a group of texts, or small corpus (Silke 2001, 2006; Neumann and Kleinmann 2013). Both approaches entail the researcher's subjective prioritisation of units of meaning which are identified as relevant to the respective study's objectives. A systematic and objective investigation of large datasets is only possible through the integration of corpus linguistic methods of discourse analysis.

Corpus linguistic analysis

Corpus linguistic analysis involves an empirical investigation of electronic texts, primarily through quantitative methods. Software is employed to query the corpus, typically with respect to the frequency and the position of, and the relationship between, particular words, or multi-word units. A subsequent qualitative analysis entails manually sifting through search results, usually in the form of concordance lines (that is, individual lines of a text which contain a given lexical item). Corpus linguistic analyses are typically employed to examine the lexico-grammatical characteristics of texts and can contribute to revealing how particular meanings are construed.

A distinction may be made between corpus-based and corpus-driven analyses. In the case of the former, the text is a 'repository of examples' (Tognini-Bonelli 2001, p. 84) which the researcher may employ to illustrate or support previously formulated claims. The text is subjectively searched through query terms selected to identify examples which serve the researcher's objectives. A corpus-driven approach, in contrast, examines the text in its entirety, and the researcher's descriptions of findings must be consistent with corpus evidence (Tognini-Bonelli 2001). As a result, the existence and absence of queried terms is of equal importance. The systematic, objective and comprehensive sifting of data is only possible through automated software analysis.

Methods

For our analysis of the three texts, we employed the following software: LIWC2015 (Linguistic Inquiry and Word Count) (LIWC n.d.) and Wmatrix (Rayson 2009). We shall briefly explain the purpose of each.

LIWC2015 was developed to identify latent emotional and psychological elements in discourse through the analysis of individual words in digitalised texts. These words are matched with the words contained in the inbuilt dictionary, which have been assigned categories reflecting a range of dimensions. The scale or weight of each category depends on the proportion of words in the text assigned to that category. A more detailed description of the functioning of LIWC is provided in Pennebaker et al. (2015), Tausczik and Pennebaker (2010) and Vergani and Bliuc (2015).

The initial analysis provides two sets of results. Firstly, we obtained results for four general categories termed *Summary Dimensions* (*Analytical Thinking*, *Clout*, *Authenticity* and *Tone*). Together, these provide an overview of the psychological characteristics of a focus text. Briefly, *Analytical Thinking* reflects language use that reveals 'formal, logical, and hierarchical thinking patterns'; *Clout* refers to the display of 'social status, confidence, or leadership'; *Authenticity* refers to honesty and is contrasted with deceptiveness; and *Tone* refers to 'positive or negative emotional dimensions', whereby a score below 50 indicates a negative tone (LIWC n.d.).

The second set of results concerns the overarching category of *Psychological Processes*, which consists of variables that correspond to psychological and emotional states. Each main variable is broken down into sub-variables. For instance, the variable *Drive* is broken down into five sub-variables. In our analysis of these two sets of results, we compared the general scores obtained in *Summary Dimensions* and the specific variables

comprising *Psychological Processes* with the scores (or ‘grand means’) provided by LIWC2015 Output information (see Pennebaker et al. 2015).

The second step of the analysis involved ascertaining whether the LIWC psychological variables we had identified as being salient were identified as significant by a second corpus linguistic software. To this end, we extracted the words identified by LIWC in each variable and matched these words with the corresponding semantic fields in Wmatrix. Wmatrix uses the semantic analysis system UCREL, widely used in automated tagging.³ To find the relevant semantic field, we checked the words contained in the UCREL semantic fields to ascertain whether they corresponded to the LIWC variable. For instance, the LIWC variable *Affect* contained words that corresponded to a range of UCREL semantic fields, which included *Emotions*, *Social Actions*, and *Psychological States and Actions*. We checked the concordance lines of words in each semantic field to ensure their use in context corresponded to the respective variable.

The third step of the analysis involved a keyword analysis of the words in each LIWC variable. We wished to establish whether the words from each LIWC variable were identified as keywords in Wmatrix. Keyness is a concept that is used to identify particular words in a text that occur with a relative frequency greater in the focus corpus than in a reference corpus.⁴ The analysis contributes to our understanding of the ‘aboutness’ of the text. We used two statistical measures, log-likelihood and log ratios, in our analysis of keywords and semantic fields.⁵ The log-likelihood score (henceforth LL) is used to establish whether there is a statistically significant difference between the frequency of words in the focus and the reference corpus. Following recommendations in the discipline (Wmatrix), the relative size of the difference is displayed by the log ratio score.⁶

Results

We will present the results for psychological dimensions, keywords and semantic fields in turn.

Psychological dimensions and attendant variables

Our first analysis was aimed at attaining an overview of the four general psychological dimensions contained in our focus texts, displayed in LIWC2015 as *Summary Dimensions*. As displayed in Table 1, results across all dimensions in the three texts were similar, albeit with some discernible differences. Greatest consistency was found in the dimensions *Analytical Thinking* and *Tone*. The high scores in the former and the low scores in the latter indicate that the texts were strong on formal, logical argumentation and they contained a pronounced negative emotional appeal. All texts scored highly on *Clout*, a

Table 1. Summary dimensions (LIWC2015).

	Analytic	Clout	Authentic	Tone
Breivik	91.35	65.26	16.95	23.19
Tarrant	77.38	77.61	32.13	13.66
Rumiyah	84.14	81.85	4.29	23.51
LIWC2015	56.34	57.95	49.17	54.22

measure of status, confidence and leadership. Whilst all texts scored low on *Authenticity* (a result commensurable with deception), the score for Rumiyah was particularly low.

Our second initial analysis concerned the three main psychological variables (*Drives*, *Personal Concerns*, and *Affect*) and the sub-variables that comprise these. We now turn to the results for each of these three main variables.⁷

Drives

The variable *Drives* refers to motives that encourage goal-oriented behaviour, and it comprises five sub-variables. *Power*, defined in terms of the relative impact and influence on others (Winter 2007), is associated with levels of aggression in extremist discourse (Smith 2008; Smith et al. 2008; Vergani and Bliuc 2015). As can be seen in Table 2, the scores for *Power* were higher in all focus corpora, when compared to the LIWC2015 means.

In the case of *Affiliation* (that is, the concern for maintaining relations with others), however, Tarrant contained a higher percentage of words typical for this variable; Breivik and Rumiyah did not display a notable concern for *Affiliation*. This result, however, is likely to be spurious. Many words that were unmatched by LIWC2015 were Arabic terms in Rumiyah and Breivik that signal affiliation, for instance, *ummah* (371.64 LL, 9.84 log ratio). Had these words been appropriately assigned, the *Affiliation* scores would have been higher.

Results for *Risk* and *Achievement* were somewhat stronger than the LIWC2015 means, particularly for Breivik and Tarrant. The results for *Reward*, however, were notably below the mean. Thus, while all texts contained content which acknowledged tension and uncertainty (*Risk*), and which articulated a certain goal orientation (*Achievement*), the prospect of personal benefit was not a salient drive; *Power*, however, was.

Personal concerns

Of the 12 sub-variables comprising *Personal Concerns*, only *Religion and Death* displayed scores higher than the LIWC2015 mean (across all focus texts) and contained words identified as significant in the keyword analysis. As is evident from Table 3, Rumiyah displayed the greatest concern for *Religion*. This is the highest score that Rumiyah obtained for any sub-variable comprising *Personal Concerns*. This result signals that religion constitutes a core ideological value in Rumiyah and is a vital component of ISIS recruitment discourse. This is consistent with the group's claim to be the only legitimate representative of Islamic belief. The importance of religion in the right-wing texts was not uniform (in Tarrant it was of marginal concern), and this illustrates the discrepancy possible in the discourse produced by groups (or individuals) of a (purportedly) similar ideological persuasion. In contrast, results for *Death* were similarly high across all focus texts, with the strongest emphasis evident in Tarrant.

Table 2. Drives (LIWC2015).

	Power	Affiliation	Risk	Achievement	Reward
Breivik	4.18	1.86	0.74	1.70	0.86
Tarrant	4.89	2.66	0.65	2.10	1.10
Rumiyah	3.43	1.48	0.56	1.34	0.92
LIWC2015	2.35	2.05	0.47	1.30	1.46

Table 3. Personal concerns (LIWC2015).

	Religion	Death
Breivik	2.23	0.49
Tarrant	0.77	0.80
Rumiyah	3.73	0.59
LIWC2015	0.28	0.16

Affect

Emotion is a vital element of oppositional discourse (i.e. extremist and nationalistic discourse). An appeal to the sentiments of the target audience is a strategy typically pursued for polarisation and recruitment purposes, and it commonly seeks to invoke a sense of ingroup attachment and victimhood, and a sense of inferiority, culpability, and difference of the ‘other’ (Baider and Constantinou 2014). The *Affect* variable score reflects the level of emotional language employed. As can be seen in Table 4, however, only in Tarrant did the overall *Affect* score exceed that of LIWC2015.

The *Affect* variable is comprised of both *Positive* and *Negative* emotions. All focus texts scored well below the LIWC2015 level of *Positive* emotion, with Rumiyah and Breivik scoring the lowest. This contrasted with the higher levels of *Negative* emotion. The *Negative* sub-variable is further broken down into *Anger*, *Anxiety* and *Sadness*. Only results for *Anger* differed markedly from the LIWC2015 mean, and these were particularly high in Tarrant. Previous research has linked *Anger* to mobilisation (Smith 2008; Baider and Constantinou 2014), that is, the endeavour to attract sympathisers or activists. The effect of this emotion may be countered by the potentially paralysing effect of content expressing *Anxiety* and *Sadness*. However, in light of the unremarkable scores across all three texts for these two sub-variables, this countering effect was absent.

In sum, all texts displayed a strong appeal to (negatively oriented) logical argumentation, exuded confident authority, and employed decidedly deceptive discourse. A pronounced *Power Drive* was identifiable across all texts, and this was accompanied by concerns for *Religion* and *Death*, and emotional content dominated by *Anger*.

Keywords and semantic fields

The second research question enquires into the keyness of words comprising each LIWC variable, and the keyness of the semantic fields that are associated with each psychological variable. This research question seeks to establish whether the words associated with each psychological variable, and the semantic fields that correspond to each variable, occur more frequently in the focus texts than they would occur in general language use.

To identify the keyness of words associated with each LIWC variable, we undertook a keyword analysis for each text using Wmatrix. The number of keywords in each text was

Table 4. Affect (LIWC2015).

	Affect	Positive	Negative	Anxiety	Anger	Sadness
Breivik	4.57	2.18	2.34	0.39	1.08	0.30
Tarrant	6.87	2.94	3.83	0.45	2.25	0.58
Rumiyah	4.95	2.40	2.54	0.28	1.40	0.30
LIWC2015	5.57	3.67	1.84	0.31	0.54	0.41

as follows: Tarrant 319, Rumiyaah 1615, and Breivik 2140. Due to the elevated number, we selected the first 100 keywords in each text and identified the corresponding psychological variable for each. Some keywords were not included in the LIWC2015 dictionaries and we will describe our treatment of these.

Some LIWC variables were excluded from this analysis. This involved variables which did not contain keywords that appeared within the first 100, or for which the associated semantic fields were not identified as key. Our analysis is thus limited to the *Drives: Power* and *Affiliation*, *Affect: Positive* and *Negative*, and *Personal Concerns: Religion* and *Death*. We shall present the results for each variable in turn.

Power (drive)

With respect to *Power*, 29 words were key in at least one of the focus texts. Five words had keyness scores above the threshold in all three corpora. These were: *destroy*, *attacks*, *fight*, *wealth*, and *strength*. In the case of many words, the LL score was extremely high, and this was often reinforced by a very high log ratio score (e.g. *attack* and *fight*). This signals that the magnitude of difference between the focus text and the reference corpus was particularly large for these words.

In Tarrant, 12 keywords were categorised under the variable *Power*. These were: *invaders* (298.21 LL, 7.32 log ratio), *destroy* (90.55 LL, 5.44 log ratio), *attack* (332.72 LL, 5.45 log ratio), *fight* (62.19 LL, 4 log ratio), *wealth* (52.72 LL, 4.53 log ratio), *strength* (47.3 LL, 3.73 log ratio), *power* (37.63 LL, 2.46 log ratio), *scum* (35.79 LL, 8.24 log ratio), *force* (46.53 LL, 3.24 log ratio), *victory* (108.42 LL, 4.37 log ratio), *political* (48.33 LL, 2.36 log ratio) and *invader* (75.66 LL, 9.24 log ratio).

Breivik contained only 10 keywords categorised under this variable: *Jihad* (1015.59 LL, 10.68 log ratio), *political* (685.21 LL, 2.12 log ratio), *resistance* (314.74 LL, 3.17 log ratio), *elites* (292.96 LL, 8.88 log ratio), *military* (353.69 LL, 1.94 log ratio), *warfare* (245.93 LL, 4.18, log ratio), *civil war* (260.38 LL, 4.61 log ratio), *war* (252.58 LL, 1.6 log ratio), *attacks* (300.72 LL, 3.79 log ratio) and *power* (215.33 LL, 1.59 log ratio). Finally, in Rumiyaah, the variable *Power* was expressed through 16 keywords: *Crusaders* (463.35 LL, 8.22 log ratio), *Jihad* (1105.32 LL, 11.42 log ratio) *Khilafah* (941.92 LL, 11.19 log ratio), *worship* (341.38 LL, 4.7 log ratio), *obedience* (284.15 LL, 8.51 log ratio), *lord* (398.96 LL, 2.66 log ratio), *slaves* (385.47 LL, 7.02 log ratio), *soldiers* (963.71 LL, 4.36 log ratio), *wealth* (608.16, LL, 4.86 log ratio), *tawaghit* (374.85 LL, 9.86, log ratio), *taghut* (349.22 LL, 9.75 log ratio), *Mujahidin* (804.16 LL, 10.96 log ratio), *Mujahid* (291.55 LL, 9.49 log ratio), *Istishadi* (278.73 LL, 9.43 log ratio), *fight* (477.04 LL, 3.81 log ratio), and *victory* (307.68 LL, 3.18 log ratio). Manual re-categorisation was needed for some words that had been unassigned or (incorrectly) automatically assigned to *Religion*. These included: *crusaders*, *Khilafah*, *tawaghit*, *istishhadi* and *mujahidin*, concepts that primarily express *Power*.

To illustrate the use of *Power* words, we identified significant collocates of highly salient words in each focus text (*invaders*, *resistance* and *war*). Example 1 displays the concordance lines containing these words and their collocations. As illustrated here, the word *invaders* is frequently used by Tarrant to refer to migrants in the European context; *resistance* is often used by Breivik in reference to a political undertaking

and recruitment; and *war* is conceived by Rumiyaḥ as involving the Caliphate against the West and its allies.

The second analysis inquired into the keyness of the semantic fields that corresponded to *Power* keywords. All nine corresponding semantic fields were key. The semantic fields *Warfare* (G3) and *Damaging and Destroying* (A1.1.2) displayed particularly high LL scores. This result corroborates our initial observation that all three texts share a high level of violent and aggressive language.

Example 1.

Tarrant (invaders)

To take revenge on the **invaders** for the hundreds of thousands of deaths caused by foreign **invaders** in European lands throughout history.

Thus, before we deal with the fertility rates, we must deal with both the **invaders** within our lands and the **invaders** that seek to enter our lands.

Breivik (resistance)

Efforts to consolidate and recruit patriotic **resistance** fighters in Western European prisons must be a prioritized task the coming decades.

A very important aspect of this non-military effort will involve indirect recruitment work for the armed European **resistance** movement.

Rumiyaḥ (war)

As the soldiers of the Khilafah continue waging **war** on the forces of kufr, ...

it is a must on every muwāḥhid to expand the scope of his jihad to include waging **war** on the kuffars wealth

Affiliation (drive)

With respect to *Affiliation*, 72 different keywords were identified: Breivik (35), Tarrant (25), Rumiyaḥ (26). Only two keywords appeared across all focus texts: *their* (LL: 1812.74, Rumiyaḥ; 73.71, Tarrant; 260.22, Breivik) and *people* (398.39 LL, Rumiyaḥ; 317.4 LL, Tarrant; 230.36 LL, Breivik). Several patterns can be noted with regard to this variable.

Firstly, third person pronouns were salient in Rumiyaḥ (*he* 883.11 LL; *his* 1125.24 LL; *him* 1503.93 LL; *them* 2936.32 LL; *they* 1267.38 LL). A survey of the concordance lines revealed the intensely interpersonal nature of the discourse, which involved actions and communication between opposed groups (the ‘true believers’ and the rest) and between people and God. Secondly, Rumiyaḥ contained numerous faith-related references, principally in Arabic, used to designate individuals as the religious outgroup, e.g. *kuffar* (762.51 LL, 10.88 log ratio) *kafir* (342.81 LL, 9.73 log ratio),⁸ *mushrikin* (592.71 LL, 10.52 log ratio), *murtadd* (509.41, LL, 10.3 log ratio) *murtaddin* (823.38 LL, 10.99 log ratio), *tawāghit/taḡhut*, *crusader/crusaders*, *rafidi* (336.4 LL, 9.7 log ratio) *rafidah* (269.12 LL; 9.38 log ratio) (see Example 2).

Example 2. Rumiya

the people of Sham – who have suffered greatly, at the hands of the Nusayri army and its allies of the **Rafidah** and Russian **Crusaders** ...

It is a spiteful **Crusader** enemy that has come with an awful plan to conquer the **Ummah** and empower the Jews.

If their **taghut** Erdogan feared for their families, he would have dispatched his despicable **kafir** troops

Similarly, many *Affiliation* keywords in Tarrant designated the outgroup, e.g. *traitors* (101.17 LL, 7.5 log ratio), *non-European* (41.17 LL, 9.24 log ratio), *immigrants* (70.1 LL, 5.81 log ratio), *anti-white* (55.92 LL, 7.92 log ratio), *NGOs* (49.4 LL, 9.5 log ratio), *mass (immigration)* (42.2 LL, 3.61 log ratio) (see Example 3).

Example 3. Tarrant

NGOs are directly involved in the genocide of the European people

Crush these **traitor** NGOs, kill their leadership

YOU ANTI-WHITE SCUM

Merkel, the mother of all things **anti-white** and anti-germanic

our enemies increase within our lands, driven by **mass immigration** and the **invaders** own higher birth rate

For several words the reference was ambiguous. Race-related references, for instance, were commonly used to refer to the ingroup (e.g. *my*, *our*, *white race*), but not exclusively, as Example 4 illustrates. *Peoples* was primarily (but not exclusively) a reference to the ingroup, and usually appeared with the pronouns *our* or *my* or *your*, or the adjectives *European*, *native*, *white* or *French*.

Example 4. Tarrant

enemies of our **race** that freely walk through our societies

The **race** of the elect suffers outrageous persecutions

to replace our **people** with a **race** of low intellect

The French **people** were often in a minority themselves

White **people** are failing to reproduce

In Breivik, the keywords classified under *Affiliation* were employed in Breivik's 'historicalisation of the present' (Bieber 2002), that is, in his reliance on myth-imbued accounts of historical figures, groups and events (dominated by a binary east/west, Christian/Muslim paradigm) to explain the alleged cultural threat in present-day Europe (see also Baider 2017). Most references to countries, nationalities and religions were used to illustrate this struggle for power and cultural dominance and inter-group relations. Nevertheless, several designations had an unequivocal ingroup reference, e.g. *Europe* (857.57 LL, 2.4 log ratio) /*European(s)* (678.62 LL, 5.94 log ratio), *knights* (603.67 LL, 9.93 log ratio) (or

PCCTs 276.98 LL, 8.8 log ratio), *justiciar knights* (323.14 LL, 9.02 log ratio) and others had a clear outgroup reference, e.g. *Marxist(s)* (529.61 LL, 5.76 log ratio), *multiculturalist* (300.06 LL, 8.92, log ratio), *traitors* (313.39 LL, 5.84 log ratio)(see Example 5).

Example 5. Breivik

The **PCCTS, Knights Templar** is a manifestation of this initial struggle, phase 1 of the Western European civil war

The **Knights Templar Europe** and Christendom has been under constant attack by **Islam** for the last 1400 years

the growing numbers of nationalists in W. **Europe** are systematically being ridiculed, silenced and persecuted by the current cultural **Marxist/multiculturalist** political establishments

it is our duty as **Europeans** to prevent the annihilation of our identities, our cultures and traditions and our nation states

These recommendations were accompanied by a deliberate, privileged influx of Arab and other **Muslim immigrants** into Europe in enormous numbers

We should ban **Muslim immigration**

Islam WILL be thrown out of Europe for a third time and the **Marxist, humanist** and globalist **traitors** who colluded with the **Ummah** will be held accountable for their treason.

Subsequently, we inquired into the keyness of the four semantic fields that corresponded to the keywords in *Affiliation*. Only one semantic field was key across all focus texts: *People* (S2). A manual examination of words comprising this field revealed the salience of gender-specific references across all texts, the majority of which were kinship terms. Of these, 14 different terms designated males, and just four referred to females. Rumiyah displayed the greatest variety of kinship terms, whilst Tarrant contained just two. The word *brothers* (347 LL, 3.5 log ratio) was particularly significant in Rumiyah, and the effect size signalled that it was 15 times more common in the focus corpus than in the reference corpus. The notion that all Muslims are joined by the bond of religious brotherhood is a well-known Islamic cultural concept. Rumiyah manipulates this concept to reinforce its claim to religious authenticity, and simultaneously, to appeal to potential recruits or sympathisers by projecting a sense of familial belonging, that is, to ISIS. The concordance lines in [Figure 1](#) illustrate the use of the word *brother(s)*.

Belonging to a Group (S5+) was key in Breivik and Tarrant. The absence of keyness for Rumiyah can be explained by the use of Arabic terms in this text for the corresponding collective words (e.g. *ummah* or *Jam'ah*). These are listed in Wmatrix in the 'unmatched' semantic field. A further classification problem was identified with the word *race*, a word used with significant frequency in Tarrant. Wmatrix miscategorised this word to *Competitive* (S7.3+). A manual survey of the concordance lines confirmed that all 25 examples corresponded to the category *Belonging to a group* (S5+). This re-categorisation strengthens the racial/ethnic preoccupation evident in the results for S5+ in Breivik and Tarrant. For instance, S5+ words in Tarrant included: *ethnic* (122.44 LL, 5.31 log ratio), *race* (89.76 LL, 4.12 log ratio), *ethnically* (74.1 LL, log ratio 10.09), *ethnicity* (57.63 LL, 9.72 log ratio), *racial* (169.54 LL, 7.56 log ratio). The frequency of these words was highly significant and the effect sizes were very large.

y of Raqqan to compete with their brothers in pursuing the pleasure of the Lo
 eminder for myself and my mujahid brothers , as an incitement to steadfastnes
 ear the de-sertion of your Muslim brothers in the various parts of the world
 atment . O my sisters , while our brothers , the mujahidin , are giving their
 unity here to re-mind our mujahid brothers and the Muslims in gen-eral to tak
 an the house of a spider . To our brothers in aqidah and iman in Europe , Ame
 Australia , and elsewhere , your brothers in your lands have absolved themse
 dows of swords . O our imprisoned brothers everywhere , by Allah we have not
 and ask Him to grant your mujahid brothers victory , steadfastness , and cons
 sies and of the opinions of their brothers , the Mutazilah , who are falsely
 , and the mushrikin out to be the brothers and allies of the Muslims . They s
 S give zakah , then they are your brothers in religion (At-Tawbah 11) . So
 , Our Lord , for-give us and our brothers who preceeded us in faith and put n
 finance the hijrah of his Muslim brothers to the wilayat of the Khilafah , o
 lectures and teaching some of his brothers the religion . He wasnt from among
 ey in order to assist some of the brothers in making hijrah to one of the are

Figure 1. Concordance lines for 'brothers' in Rumiyah (Wmatrix).

In sum, we believe that importance of the semantic field *Belonging to a group* (S5+) was greater than the results indicate, and that while Rumiyah conceived group belonging primarily in religious and geographic terms, in Breivik and particularly in Tarrant, the racial/ethnic element was more salient.

Death (personal concerns)

The keyword analysis uncovered five words in the variable *Death* within the top 100 keywords in each text: *death* (46.02, LL; 2.58, log ratio) and *kill* (75.07, LL; 4.26, log ratio) (in Tarrant), *killed* (912.8 LL, 4.32 log ratio) and *killing* (542.79 LL, 4.54 log ratio) (in Rumiyah) and *genocide* (316.04 LL, 8.99 log ratio) (in Breivik).

In Tarrant the keyword *death* collocated with *accept*, *embrace* and *traitor*. *Embrace* and *accept* were used by Tarrant to call upon his fellow white (ingroup) to 'accept death' when acting to protect their race, while the word *traitor* was used to refer to the punishment of those (particularly 'high profile enemies') who he perceived as acting against the interests of the white race. The significant collocates of *kill* were *invaders*, named politicians (e.g. *Merkel*, *Sadiq Khan*) and 'local traitors' (*anti-whites*, *drug dealers*) (see Example 6).

Example 6. Tarrant

They will soon find out how wrong they truly are. **Traitors** deserve a **traitors death**. No matter if it takes 3 years or 30 years, these people must pay for their disgusting attacks upon our race.

The Danger of the Invader If you were **to kill sixty armed invaders** having shown the will and the intent to bring harm to your nation and people, you would be hailed a hero, given your ...

What better sign of the white rebirth than the removal of this invader? **KILL ANGELA MERKEL, KILL ERDOGAN, KILL SADIQ KHAN** The Paradox of the diverse equality

In Breivik, the most significant collocates of the word *genocide* were *Armenian*, *demographical*, and *cultural*. In the case of Breivik and Rumiyah, the focus on death was often embedded in accounts of historical events, that is, the perceived present imperative for violent action was again justified by invoking a mythical historical context. The Qu'ran

was a commonly cited source to justify death in Rumiyaah, while in Breivik it was used to associate Islam with death (see Example 7).

Example 7. Breivik

Stop the ongoing European **cultural and demographical genocide** facilitated by the cultural Marxists/multiculturalists, suicidal humanists, and capitalist globalist elites

Our battle on the other hand involves Cultural Conservatism, our duty and right to resist **Cultural genocide** and **Islamic demographic warfare**.

Being thrown off cliffs, burned alive, or drowned in rivers. During the **Armenian Genocide**, the Turkish countryside became littered with decomposing corpses.

In Rumiyaah the keyword *killed* collocated with *wounded*, *murtaddin*, and *others* while the word *killing* collocated with *injuring*, *wounding*, and *succeeded*. Both keywords were used in contexts of reporting the killing of enemies (often quantified), a strategy used by Rumiyaah to project the image of a victorious state (see Example 8).

Example 8. Rumiyaah

near Ittihad University, where fierce clashes for several hours involving various weapons, resulting in **30 murtaddin** being **killed** and **others wounded**.

They succeeded in **killing** and **injuring** nearly **60 murtaddin** before attaining shahadah.

The corresponding semantic field *Dead (L1-)* was key across all datasets, a result that was consistent with LIWC2015 Death. The semantic field *Alive (L1+)* also corresponded to LIWC2015 *Death*, and this field was also key across all texts; very few words comprised this semantic field, however, primarily *life*, *lives*, and *alive*. Nevertheless, an examination of the concordance lines revealed that these words were often used to refer to death (see Example 9). We therefore surmise that the results for *Dead (L1-)* should have been higher.

Example 9.

the thousands of European **lives** lost to terror attacks throughout Europe [Tarrant]

six years old Greek boy was burned **alive** by the Turks [Breivik]

attempts to kill him by burning him **alive** [Rumiyaah]

Religion (personal concerns)

A total of 61 words received significant LL scores for keyness in at least one of the three texts. Of these, 32 keywords related to Islam or Muslims, and many of these had very high LL and Log ratio scores. Breivik contained the highest number (28 keywords), followed by Rumiyaah (10 keywords). The most salient words in both texts were *Islam* (2360.78 LL, 8.13 log ratio [Breivik], 1217.09 LL, 7.83 log ratio [Rumiyaah]), *Islamic* (2039.07 LL, 6.63 log ratio [Breivik]; 902.01 LL, 6.15 log ratio [Rumiyaah]), *Muslims* (2788.87 LL, 7.22 log ratio [Breivik]; 1461.95 LL, 7.34 log ratio [Rumiyaah]) and *Jihad*.

A survey of the concordance lines containing these keywords revealed a strong focus in Breivik on the so-called Islamic colonisation and Islamisation of the west, and this theme

was accompanied by references to historical events which were interpreted to support his narrative of the threat of Islam to European civilisation.

Tarrant, in contrast, did not contain religious references in the 100-keyword list apart from the word *beliefs* (79.82 LL, 5.73 log ratio). Further review of the keywords revealed only few keywords overall with religious associations: *Islamophobe* 16.47 LL, 7.92 log ratio; *mosque* 23.47 LL; 6.33 log ratio; *mosques* 27.96 LL; 7.92 log ratio; *Islamic* 23 LL, 4.33 log ratio, *Muslim* 18.64 LL, 4.24 log ratio; *Muslims* 15.33 LL, 4.33 log ratio; *minarets* 16.47 LL, 7.92 log ratio. A survey of the concordance lines revealed that these words were used on relation to Tarrant's plan of assault, which the author justifies due to the greater likely impact of a Muslim rather than an alternative religious minority target (e.g. Jewish or Sikh), and Muslims allegedly higher fertility rates. His focus on the Islamic building was thus not motivated by a concern for religion as much as a concern for the migrant minority groups that frequent these locations in Europe.

In light of this discrepancy between the salience of religion in Breivik and Tarrant, we conducted a further analysis of the dispersion (or distribution) of keywords related to religion in each text.⁹ With values ranging from 0 to 1, dispersion identifies the relative spread of focus words throughout a text. Values closer to 0 signal a skewed occurrence of the word, and values closer to 1 represent even distribution. As can be seen in Table 5, in Breivik the normalised frequency of keywords (i.e. per 1000 words) related to religion was higher and the dispersion value was closer to 1. This signals greater evenness or uniformity of attention to religion in Breivik.

In Tarrant, however, the only religion word with a relative frequency of occurrence higher than in Breivik, *mosque*, displayed a very skewed dispersion. The concordance lines confirmed the aforementioned finding that it was almost exclusively related to logistical rather than ideological deliberations. While the frequency and dispersion levels were somewhat higher for *Islam* and *Muslim* than *mosque*, these were nevertheless considerably lower than the levels for words related to race (these were keywords in the *Affiliation* variable). Tarrant employed race-related words more frequently and with greater regularity throughout the text. In contrast, whilst Breivik focused on both religion and race throughout the text, the prominence of the words *Islam*, *Muslim*, *Christian*, and *Europe* indicate that his primary concern was religion and civilisation.

Table 5. Relative frequency and dispersion of keywords (religion and race).

Terms	Tarrant Norm. freq.	Dispersion ratio	Breivik Norm. freq	Dispersion ratio
Islam*	0.60	0.553	4.35	0.748
Muslim*	0.60	0.328	4.71	0.740
Sharia	0.06	0.000	0.31	0.588
mosque*	0.54	0.241	0.13	0.607
jihad*	–	–	1.30	0.699
Quran	–	–	0.18	0.329
Prophet	–	–	0.21	0.396
religion	0.12	0.000	0.58	0.687
Christian*	0.42	0.600	2.88	0.752
crusader*	–	–	0.14	0.619
Race	1.63	0.815	0.18	0.686
Europe*	6.16	0.814	5.53	0.846
White	2.11	0.720	0.26	0.779
traitor*	1.27	0.613	0.36	0.594
Immigra*	1.99	0.716	0.90	0.686

*The asterisk indicates any letters might follow. For instance, *Islam** includes *Islamic* and *Islamist*.

In our examination of the semantic fields that related to *Religion*, the results were consistent with LIWC2015. Whilst religion-related content in Breivik and Rumiya corresponded to a wide range of semantic fields, and the LL scores for these were usually very high, only one significant semantic field could be identified in Tarrant.

Affect

Our keyword analysis of words comprising the semantic fields that corresponded to the *Affect* variable found no keywords associated with positive emotions in any text apart from the word *worship* (341.38 LL, 4.7 log ratio) in Rumiya. Most *Affect* keywords were associated with *Anger*. A considerable number were also categorised as keywords under *Power* (e.g. *attack, fight, destroy*) and some were also categorised under *Death* (e.g. *kill, killing*).

Our examination of the corresponding semantic fields for the words comprising the *Affect* variable identified 19 corresponding semantic fields that were key for at least one of the three focus texts. Two semantic fields were significant across all texts: *Violent/angry* (E3-), and *Formal/unfriendly* (S1.21), and the respective effect sizes signalled that these semantic fields were around three and five times more frequent in the focus texts than the reference corpus. The semantic field analysis revealed the range of emotions (both positive and negative) identified as key, a factor particularly apparent in Breivik. In contrast, in Tarrant only two semantic fields that corresponded to negative emotions were key, but no positive emotion. This difference was not apparent in the LIWC2015 results.

Discussion

Our analysis of three prominent extremist texts sought to identify the extent to which ideologically opposed extremist texts display similarities with respect to the psychological and semantic content. At a conceptual level, our analysis revealed strong similarities across all texts with respect to the three dimensions previously discussed in Holbrook (2013): convergence of issues, emotive language and strategies, and the use of force.

With respect to convergence of issues, the analysis revealed a preoccupation across all texts with the delineation of social groups and with inherent inter-group incompatibility. A pervasive theme in our datasets concerned identity politics. This was manifested by a preoccupation with the assigning of group identities and the creation of essentialist well-defined boundaries between categories. Groups were portrayed as being in a historically situated struggle over territory, authority or power and resources. The ideational construction of inherent antagonism between groups was embedded in broader narratives of historical injustice, bellicose relations and intergroup incompatibility, which invested contemporary narratives with a mythical dimension (Bieber 2002; Baider 2017). Such linkages facilitate the perception of logical reasoning, as such themes have precedents in enduring historical narratives concerning civilisational exceptionalism (whether European, Christian, or Islamic) (see, for instance, Zúquete 2008).

Identity politics relies on epithets to distinguish groups and their constituents. The array of epithets used proved to be richer for outgroups. These were identified through references to nationality, religion, residence status, skin colour, political orientation, or

social concern. The particularly broad range of epithets in Tarrant and Breivik reflected the high level of disaffection with social change in the early twenty-first century, which has dislodged the white European male from traditional positions of uncontested power. Religion and cultural/civilisational delineations dominated the discursive construction of groups in Rumiyah. The more limited array of epithets to denote the ingroup in all texts reflected the desired portrayal of religious or ethnic/cultural authenticity and homogeneity. Of interest was the intensity of anger and hate expressed by epithets to denote individuals considered 'traitors'. All texts are explicit with regard to the identity of 'traitor' groups or individuals. Breivik creates categories of traitors and outlines in detail their impending punishment; Tarrant emits imperatives commencing with *KILL*; and Rumiyah specifies their nationalities or their names (usually military and political leaders) with information about their death or alleged depravity. This vilification of individuals from the cultural ingroup or religion who are unsympathetic to extremist ideology has been widely documented in social media, and can be associated with 'off-line' threatening and violent actions (e.g. the murder of German politician Walter Lübke in June 2019). Previous research has also noted the importance of epithets in combination with words to denote (usually) the outgroup (e.g. *kuffar*, *migrants* etc.) in automatic detections of online hate speech (De Smedt et al. 2018).

The high levels of the variable *Clout* in all texts were indicative of the authoritative and statusful position assumed by the authors in performing the role of distinguishing 'rightful' behaviours and beliefs of the respective ingroup. This claim to centrality is a mobilisation strategy shared among ideological extremists, and the confidence in their claim to represent mainstream views is bolstered by the indirect (or even explicit) sanctioning of such views among political and media elites (Pratt 2015).

Explicative 'stories' (textual and visual) employed by media, politicians and influencers matter (Ebner 2018). Current narratives construed by some media outlets and politicians in the US and Europe that frame minorities and immigrants as a social, economic or political menace creates a context in which assigning individuals to negatively framed categories, and referring to these with censorious or alarmistic epithets, creates an influential precedent for discriminatory groupist-centred explications. For instance, associating Mexican immigrants with rape (US President Trump in 2018), and urging democrat politicians to return to their 'broken and crime infested places' (Trump in July 2019) contributes to an interpretive schema capable of rationalising the massacre at El Paso by Patrick Crusius in August 2019. In the New Zealand context, public media have afforded figures such as Paul Henry and Brian Tamaki a powerful platform for their often bigoted and inflammatory narratives, and have only since the Christchurch massacre demonstrated greater cognisance of the resultant harm of 'casual racism' (see, Saxton 2019).

Identity politics relies on the emotional element (i.e. *Affect*) in discourse. Dominant themes in extremist discourse (typically related to cultural, social, or economic rivalry between social groups, or impending peril) may foment anxiety. However, when accompanied by the element of anger, a dearth of positive emotion, combined by high levels of language related to power, the content is construed in a manner conducive to inciting a response entailing antagonism, violence, or rage (considered mobilising factors) (Baider and Constantinou 2014; Ebner 2018). This was consistent with our results. All texts employed a divisive binary discourse that focused on the victimhood

of the ingroup and the imperative for defensive action and, on the other hand, construed the outgroup as a brutal adversary. The particular grievances employed as injustices centred on the expansion or domination of a particular religion (or religious interpretation), or cultural identity. Control over resources (political, economic, and territorial) was linked to these identities. Whilst for Rumiya and Breivik the religious element was a central thematic grievance (and motive to incite anger), in Tarrant the vulnerability of (European, white) cultural identity was underscored.

An important finding in this study constitutes the distinction between religious and racially framed grievances. Tarrant states explicitly that his primary concern is demographic replacement, and this was confirmed by our analysis, which identified a relatively low level of anti-Islamic (or Islamophobic) content. According to our findings, Islamophobic convictions and religion were not the primary motivators for the Christchurch massacre, but rather the increased presence of ethnic minority migrants and their allegedly greater reproductive proclivity. This insight is important in light of the immediate assumption that a religion was targeted, as was evident from the official government response, and the academic response to date.¹⁰ This finding also speaks to the, at times, contested conceptual distinction between Islamophobia and racism (see Pratt and Woodlock 2016), and supports the view that Muslims can be the target of discriminatory acts that are primarily motivated by racism rather than religious intolerance.

Several methodological shortcomings require acknowledgement. The software, Wmatrix and LIWC2015, were unable to accommodate easily the type of language employed in our texts. As noted in De Smedt et al. (2018), language employed in extremist texts and Web 2.0 platforms displays many non-standard features (e.g. symbols, word creations, and inconsistent spellings). Moreover, Islamic extremist discourse in English relies heavily on Arabic words to convey particular meanings. The resultant blended English-Arabic discourse is experienced as a single whole, rather than as alternating codes, in a manner consistent with translanguaging (García and Wei 2013).

These linguistic features were assigned to the *Unmatched* (Z99) category by Wmatrix. Due to the high proportion of Arabic words, this was the most statistically significant semantic field in Rumiya and was also of importance in Breivik. Furthermore, many Arabic words had been categorised under the semantic field *Religion* (S9) when their meaning actually corresponded to a different category, e.g. *Power* (S7.1+) or *Belonging to a Group* (S5+). The mis-categorisation also entailed instances where the semantic meaning did not align with the pragmatic meaning (e.g. some words in *Alive* L1+ field should have been in *Dead* L1-), and this affected the level of keyness of some semantic fields.

With respect to LIWC2015, some keywords which were identified as highly significant by Wmatrix, such as *diversity* (103.21, LL) and *ethnic* (122.44, LL) in Tarrant, *knights* (543.41, LL) and *multiculturalist* (339.12, LL) in Breivik, were not categorised under any LIWC2015 variable, indicating that these words were not foreseen in the inbuilt dictionary. Furthermore, some keywords were not classified under all relevant LIWC2015 variables; for example, the word *nations* (153.54, LL) in Tarrant was confusingly classified under *Motion* but not under *Affiliation*. This affected Rumiya particularly due to the number of Arabic words classified solely under *Religion*; for instance, *mujahidin* (804.16 LL) and *Tawaghit* (374.85, LL) could have equally been categorised under *Power*.

To overcome this issue, we manually re-categorised the first 100 keywords to the correct Wmatrix semantic fields and then grouped these under the corresponding LIWC2015 variable. This enabled a more realistic insight into the type and the significance of the words used to convey particular concepts in each text. We consider the incorporation of two different corpus linguistic tools (each with its respective inbuilt reference corpus) as an important measure to strengthen reliability. We recommend software developers to update their inbuilt corpora and review the validity of their automatised categorisations, or alternatively enable users to adapt automatised functions to their specific needs.

As is characteristic of many corpus linguistic studies, our specific focus was on the language used to construct the text at lexical and semantic levels. Through the software employed, we were able to identify explicit but not implicit meanings (such as metaphoric language). As such, the texts may have contained additional meanings that we neglected. A final limitation concerns the parameters of our study. Our analysis focused solely on text and we are unable to verify whether the meanings we identified in the text, and the strategies used to express these meanings, accord with the respective authors' intentions. Nor are we able to shed light on how these texts were understood within the milieu of the like-minded. While our study is responsive to the call for the use of primary data in research into extremist groups (e.g. Schuurman and Eijkman 2013), the need for ethnographic fieldwork with an emic perspective remains acute (see Toros 2008).

Glossary

Pharaoh	This term is used to refer to the current political leaders of Islamic countries who are perceived as tyrants who refuse to submit and conform to the will of Allah (as interpreted by ISIS).
Hypocrites or the munafiqun (plural), munāfiq (singular)	This term is used to refer to people who (misleadingly) claim to be Muslims and who act with the intention of harming the Muslim community. In ISIS discourse, the term is appropriated to refer to Muslims, whether living inside or outside ISIS territories, who do not accept the ISIS interpretation of Islam, and who fight against ISIS.
Istishadi	Suicide martyr
Jahiliya	This term refers to the time in Arabia prior to Islam and can be translated as 'the age of ignorance'. This term is contemporised in ISIS discourse to refer the current state of the Muslim Ummah and their alleged distance from their faith.
Jam'ah	The Muslim community
Jihad	In Islam, <i>Jihad</i> is divided into two categories: the greater and the lesser <i>Jihad</i> . The greater (internal) <i>Jihad</i> refers to the struggle within oneself to fight evil and worldly temptations. The lesser (external) <i>Jihad</i> involves fighting against enemies who attack Islam, Muslim countries or communities.
Khilafah	Islamic state
Kuffar (plural), kafir (singular)	Disbelievers
Mujahidin, (plural), mujahid, (singular)	Fighters or soldiers
Murtaddin (plural), murtadd (singular)	A term used to refer to someone who no longer believes in Islam, equivalent to apostate.
Mushrikin (plural)	Polytheists
Rafidi (singular), rafidah (sect)	Rafidah comes from the verb Rafad, which means to reject. This name is given to Shi'a because they reject the first three of the four al-Rāshidūn, the first four caliphs (Khalifa) after Muḥammad, because they claim that the fourth Caliph 'Alī, should have succeeded in the first place being Prophet Mohammad's son in law.
Tawaghit (plural), taghut (singular)	Tyrants, dictators
Ummah	The direct translation of the term is the nation, and is used in Islamic literature to refer to the Muslim community as a whole.

Notes

1. Radicalisation is considered a process of gradually ‘increasing extremity of beliefs, feelings and behaviours in support of intergroup conflict and violence’ (McCauley and Moskaleiko 2008, p. 415).
2. We shall henceforth refer to each text as Tarrant, Breivik and Rumiyah.
3. This is further explained here: <http://ucrel.lancs.ac.uk/usas/>.
4. In most cases (including this study), the reference corpus is a general corpus, in that it contains a wide variety of texts and topics. Wmatrix uses the British National Corpus (BNC) Sampler as an inbuilt reference corpus.
5. The threshold value for keyness when using the log-likelihood statistic is 15.13 (this corresponds to a p -value of $p < .0001$). While this threshold may be used in linguistics, a lower threshold of 3.84 ($p < .05$) may be used in the social sciences (see Wmatrix).
6. Each increment of the ratio score represents a doubling of the size of difference. If a word has the ratio of 1, it is two times more common in the focus corpus than in the reference corpus, a ratio of 2 if it is four times more common, a ratio of 3 if it is eight times more common, etc.
7. Space restrictions preclude a discussion of all results.
8. The two forms of the same word reflect alternative spellings or express the singular and plural form; both forms had high keyness scores.
9. This was conducted using Wordsmith Tools (Scott 2016).
10. See, for instance, the special issue of the New Zealand Journal of Psychology, 2019, 48(1).

Acknowledgements

The New Zealand Office of Film and Literature Classification has classified two texts used in this study, *The Great Replacement* and the 13 issues of *Rumiyah*, as ‘objectionable material’. It is against the law to be in possession of these texts in New Zealand. We gratefully acknowledge the exemption granted to us by this Office for the purpose of undertaking this study.

Disclosure statement

No potential conflict of interest was reported by the authors .

ORCID

Louisa Buckingham  <http://orcid.org/0000-0001-9423-0664>

Nusiebah Alali  <http://orcid.org/0000-0003-3170-170X>

References

- Baider F. 2017. Thinking globally, acting locally: mainstream supremacist concepts within a local socio-historical context. *Journal of Aggression Language and Conflict*. 5(2):178–204.
- Baider FH, Constantinou M. 2014. How to make people feel good when wishing hell: golden dawn and national front discourse, emotions and argumentation. In: J. Romero-Trillo, editor. *Yearbook of corpus linguistics and pragmatics 2*. Cham: Springer; p. 179–210.
- Baker P, Gabrielatos C, Khosravini M, Krzyżanowski M, McEnery T, Wodak R. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*. 19(3):273–306.
- Berntzen L. E., Sandberg S. 2014. The collective nature of lone wolf terrorism: Anders Behring Breivik and the anti-Islamic social movement. *Terrorism and Political Violence*. 26(5):759–779.
- Bieber F. 2002. Nationalist mobilization and stories of Serb suffering: The Kosovo myth from 600th anniversary to the present. *Rethinking History*. 6(1):95–110.

- Conway M. 2017. Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Studies in Conflict & Terrorism*. 40(1):77–98.
- Costello M, Hawdon J. 2018. Who are the online extremists among us? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials. *Violence and Gender*. 5(1):55–60.
- Cottee S, Cunliffe J. 2018. Watching ISIS: how young adults engage with official English-language ISIS videos. *Studies in Conflict & Terrorism*. 1–25. doi:10.1080/1057610X.2018.1444955.
- Cottee S, Hayward K. 2011. Terrorist (e) motives: the existential attractions of terrorism. *Studies in Conflict & Terrorism*. 34(12):963–986.
- De Smedt T, De Pauw G, Van Ostaeyen P. 2018. Automatic detection of online jihadist hate speech. CLIPS Technical Report Series. 7:1–31. [accessed 2009 November 10]. <https://arxiv.org/ftp/arxiv/papers/1803/1803.04596.pdf>.
- 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.
- Ebner J. 2018. *The rage: the vicious circle of Islamist and far-right extremism*. London: I.B. Tauris.
- García O, Wei L. 2013. *Translanguaging: language, bilingualism and education*. New York (NY): Palgrave Macmillan.
- Gilbert J. 2013. *Patched: the history of gangs in New Zealand*. Auckland: Auckland University Press.
- Halverson J, Corman S, Goodall HL. 2011. *Master narratives of Islamist extremism*. Cham: Springer.
- Hawdon J, Costello M, Barrett-Fox R, Bernatzky C. 2019. The perpetuation of online hate: a criminological analysis of factors associated with participating in an online attack. *Journal of Hate Studies*. 15(1):157–181.
- Holbrook D. 2013. Far right and Islamist extremist discourses: shifting patterns of enmity. In: Taylor M, Currie PM, Holbrook D, editors. *Extreme right wing political violence and terrorism*. New York: Bloomsbury Publishing; p. 215–237.
- Lehr P. 2013. Still blind in the right eye? A comparison of German responses to political violence from the extreme left and the extreme right. In: Taylor M, Currie PM, Holbrook D, editors. *Extreme right wing political violence and terrorism*. London: New York: Bloomsbury Publishing; p. 187–214.
- Liebman CS. 1983. Extremism as a religious norm. *Journal for the Scientific Study of Religion*. 22:75–86.
- LIWC [Linguistic Inquiry and Word Count]. <http://liwc.wpengine.com/interpreting-liwc-output/>.
- Mahood S, Rane H. 2017. Islamist narratives in ISIS recruitment propaganda. *The Journal of International Communication*. 23(1):15–35.
- McCauley C, Moskalenko S. 2008. Mechanisms of political radicalization: pathways toward terrorism. *Terrorism and Political Violence*. 20(3):415–433.
- Neumann P, Kleinmann S. 2013. How rigorous is radicalization research? *Democracy and Security*. 9(4):360–382.
- O'Hara K, Stevens D. 2015. Echo chambers and online radicalism: assessing the internet's complicity in violent extremism. *Policy & Internet*. 7(4):401–422.
- Pennebaker JW, Boyd RL, Jordan K, Blackburn K. 2015. *The development and psychometric properties of LIWC2015*. Austin: University of Texas at Austin.
- Pratt D. 2015. Islamophobia as reactive co-radicalization. *Islam and Christian–Muslim Relations*. 26(2):205–218.
- Pratt D., Woodlock R. 2016. Introduction: understanding Islamophobia. In: Pratt D., Woodlock R., editors. *Fear of Muslims?*. Cham: Springer; p. 1–18.
- Rayson P. 2009. *Wmatrix: a web-based corpus processing environment [computer software]*. Lancaster: Lancaster University.
- Saxton A. 2019. 'Go back to your country': community rallies against islamophobes. <https://www.stuff.co.nz/national/113514901/go-back-to-your-country-community-rallies-against-islamophobes>.
- Schuurman B, Eijkman Q. 2013. Moving terrorism research forward: the crucial role of primary sources. *ICCT Background Note*, 1–11.

- Scott M. 2016. Wordsmith tools version 7. Stroud: Lexical Analysis Software.
- Silke A. 2001. The devil you know: continuing problems with research on terrorism. *Terrorism and Political Violence*. 13(4):1–14.
- Silke A. 2006. The impact of 9/11 on research on terrorism. In: Ranstorp M, editor. *Mapping terrorism research*. Abingdon: Routledge; p. 90–107.
- Smith AG. 2008. The implicit motives of terrorist groups: how the needs for affiliation and power translate into death and destruction. *Political Psychology*. 29(1):55–75.
- Smith AG, Suedfeld P, Conway LG, Winter DG. 2008. The language of violence: distinguishing terrorist from nonterrorist groups by thematic content analysis. *Dynamics of Asymmetric Conflict*. 1(2):142–163.
- Tausczik YR, Pennebaker JW. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*. 29(1):24–54.
- Tognini-Bonelli E. 2001. *Corpus linguistics at work*. Benjamins: Amsterdam.
- Toros H. 2008. Terrorists, scholars and ordinary people: confronting terrorism studies with field experiences. *Critical Studies on Terrorism*. 1(2):279–292.
- Van der Valk I. 2013. Youth engagement in right wing extremism: comparative cases from the Netherlands. In Taylor M, Currie PM, Holbrook D, editors. *Extreme right wing political violence and terrorism*. New York: Bloomsbury Publishing. p. 129–149 .
- Vergani M, Bliuc AM. 2015. The evolution of the ISIS' language: a quantitative analysis of the language of the first year of Dabiq magazine. *Security, Terrorism and Society*. 2(2):7–20.
- Winter D. G. 2007. The role of motivation, responsibility, and integrative complexity in crisis escalation: comparative studies of war and peace crises. *Journal of Personality and Social Psychology*. 92(5):920.
- Wmatrix [Corpus Analysis and Comparison Tool]. [accessed 2019 July 18]. <http://ucrel.lancs.ac.uk/wmatrix/>.
- Zúquete JP. 2008. The European extreme-right and Islam: new directions? *Journal of Political Ideologies*. 13(3):321–344.