SN Soc Sci (2021) 1:238
https://doi.org/10.1007/s43545-021-00240-4

ORIGINAL PAPER

Dehumanisation of 'Outgroups' on Facebook and Twitter: towards a framework for assessing online hate organisations and actors

Mohamad Abdalla[1,2] · **Mustafa Ally**[3] · **Rita Jabri-Markwell**[4]

Received: 2 January 2021 / Accepted: 17 August 2021 / Published online: 22 September 2021

Abstract

Whilst preventing dehumanization of outgroups is a widely accepted goal in the field of countering violent extremism, current algorithms by social media platforms are focused on detecting individual samples through explicit language. This study tests whether explicit dehumanising language directed at Muslims is detected by tools of Facebook and Twitter; and further, whether the presence of explicit dehumanising terms is necessary to successfully dehumanise 'the other'—in this case, Muslims. Answering both these questions in the negative, this analysis extracts universally useful analytical tools that could be used together to consistently and competently assess actors using dehumanisation as a measure, even where that dehumanisation is cumulative and grounded in discourse, rather than explicit language. The output of one prolific actor identified by researchers as an anti-Muslim hate organisation, and four (4) other anti-Muslim actors, are discursively analysed, and impacts considered through the comments they elicit. Whilst this study focuses on material gathered with respect to anti-Muslim discourses, the findings are relevant to a range of contexts where groups are dehumanised on the basis of race or other protected attribute. This study suggests it is possible to predict aggregate harm by specific actors from a range of samples of borderline content that each might be difficult to discern as harmful individually.

Keywords Dehumanisation · Out-groups · Dangerous organisations · Digital platform policy · Right wing extremism · Content moderation

✉ Mohamad Abdalla
Mohamad.Abdalla@unisa.edu.au

Extended author information available on the last page of the article

Introduction

Increasingly, researchers are analysing the ecosystems that socialise individuals towards extremist violence. Researchers from Macquarie and Victoria Universities have published the first study mapping the online activity of right-wing extremists (RWE) in New South Wales. Significantly, their research found that dehumanisation existed on 'low-risk' platforms like Facebook and Twitter 'without violating platform moderation policies' (3).

Whilst there is growing recognition of the need to address the online environments that socialise individuals towards violence, research is missing on how and when interventions should be made.

Thus, the starting point for this study was not to map the activity of certain actors, but to examine a distinct action and harm: dehumanisation. We chose this approach acknowledging that definitions of extremism and terrorism are contested, and therefore policy responses depending on those definitions can be fragile. Where we do refer to right-wing extremist, we use the definition from the abovementioned study, that being 'individuals, groups, and ideologies that reject the principles of democracy for all and demand a commitment to dehumanising and/or hostile actions against out-groups' (Department of Security Studies and Criminology, 2020, p. 1).

Referring to the Australian terrorist who carried out the Christchurch attack, Lentini (2019, p. 43) explains that,

> Tarrant's solution to the crisis – indeed one on which he felt compelled to enact – was to annihilate his enemies (read Muslim migrants). This included targeting non-combatants. In one point in his 'manifesto', he indicates that they constitute a much greater threat to the future of Western societies than terrorists and combatants. Thus, he argues that it is also necessary to kill children to ensure that the enemy line will not continue…Tarrant indicated that, when trying to remove a nest of snakes, the young ones had to be eradicated. Regrettably, children were amongst those whom he allegedly shot and killed.

The nest of snakes or vipers is a metaphor to describe Muslims that continues to find a home on mainstream social media. Dehumanisation takes violent and vile ideas into the realm of proper, necessary action (Maynard and Benesch 2016). It does this by disfiguring or erasing the humanity of the victim group.

Analysing the anti-Muslim field is broadly purposeful because it is considered to be a gateway to 'gradually introducing more racially and politically extremist messages to a large audience of potential supporters' (Peucker et al. 2018).

Canadian (Davey et al. 2020), Australian (Peucker et al. 2018, p. 7), US (Institute of Strategic Dialogue 2020), and UK (Allchorn and Dafnos 2020) research has found Muslims to be a favoured 'out-group' around which radical right-wing activism or extremism coalesces. Of increasing concern is that the 'highly volatile nature' of the far right milieu means that escalation from extremist thinking to action is not uncommon (Peucker 2020). Anti-Muslim hate organisations are also

more able to raise funds compared to white supremacist or nationalist organisations publicly (Institute for Strategic Dialogue and Global Disinformation Index 2020). Evidently, this fundraising also brings profit to digital platforms.

Facebook and Twitter policy

Following the Christchurch massacre, Facebook announced that it would ban praise, support, and representation of white nationalism and separatism on Facebook and Instagram. However, its own civil rights audit (Murphy 2020) found,

> [T]he policy is too narrow in that it only prohibits content expressly using the phrase(s) "white nationalism" or "white separatism," and does not prohibit content that explicitly espouses the very same ideology without using those exact phrases…[T]his recommendation must be prioritized (49–50).
> By all reports, Facebook hasn't yet grappled with the question of how to identify white nationalist ideology where it isn't overtly named.

> Facebook claims to use a combination of AI and human expertise to remove content praising or supporting *any organisation* on their 'list', which includes 250 white supremacist organisations (51) since Christchurch.

> Expanding on the category of 'terrorist propaganda' which typically relies on formalised branding and external designation and proscription lists, Facebook has created the category of 'hate organization'. Once meeting a threshold for an organization, Facebook investigates whether it has 'an ideology, statements, or physical actions that attack individuals based on a protected characteristic' (Facebook 2020).

> Acknowledging that 'hate organisation' didn't capture ideologically connected movements, in October 2020, Facebook moved to prohibit 'Violence-Inducing Conspiracy Networks, such as QAnon'. Facebook reportedly also has internal metrics for determining how inciteful of violence a particular conspiracy network may be (Mac and Silverman 2020).

> Twitter expressly bans 'violent organisations'. Twitter (2020) does not allow anyone to 'promote violence against or directly attack or threaten other people based on a range of protected characteristics. However, their threshold for disallowing accounts is an account 'whose primary purpose is *inciting harm* towards others on the basis of these categories' [emphasis added].

The last vestige of protection against the spread of dehumanising conspiracy theories like the Great Replacement narrative, are their policies on hate speech (ISD and GDI 2020, p. 5). These policies typically involve the assessment of individual pieces of evidence.

For example, Facebook's policy prohibits direct attacks on the basis of protected characteristics, which include a person's 'religious affiliation', and defines 'attack' as "violent or dehumanizing speech, harmful stereotypes, statements of inferiority, or calls for exclusion or segregation." Their policy also does not allow 'Dehumanizing

speech or imagery in the form of comparisons, generalizations, or unqualified behavioural statements (in written or visual form)', and links it to a range of classically dehumanising comparisons such as 'insects… filth, bacteria, disease… sub-humanity… violent and sexual criminals'.

Twitter's hateful conduct policy prohibits the dehumanisation of people on the basis of religion. As of late July 2020, it also prohibited third party malicious links that breach its hateful conduct rules.

Facebook (and by relation Instagram), Twitter, Youtube and LinkedIn recognize dehumanization as a particularly dangerous form of hatred as it removes moral objections one may have to enact violence, even mass violence, against women (Marczak 2018), children (Lentini 2019), and civilians more broadly within a target group.

Aims of this research project

By *explicit dehumanisation*, we refer to classically dehumanising terms such as comparing a human group to animals, bacteria, filth, disease, weeds, subhuman beings, inanimate (non-living) objects or supernatural creatures. Both Facebook and Twitter appeared to struggle to detect explicit dehumanisation in their comment threads. Our query thus turned to the material that precipitated explicit dehumanisation in the comment threads: Did it have dehumanising properties?

Therefore, this study aimed to answer three questions: (1) whether explicit dehumanising language directed at Muslims is detected by auto-detection or content review tools on Facebook and Twitter; (2) whether explicit dehumanising language was needed to successfully dehumanise the 'other'—in this case, Muslims; and finally (3) what were the characteristics to dehumanising discourse.

Theoretical framework for identifying dehumanisation

This research is grounded in the understanding that the specificities of language, and discourse more broadly is powerful in the propagation of extreme ideas (Wodak 2015).

This study builds on Maynard and Benesch's conclusion that dehumanisation can be carried out without 'hatred or blatant exclusionary discourse' (70), by analysing the operation and effectiveness of different modes of dehumanisation online.

According to the authors, dehumanisation is the most frequently employed technique in dangerous speech, where '[t]argets … are described in a variety of ways that deny or diminish their humanity, reducing the moral significance of their future deaths or the duties owed to them by potential perpetrators' (80).Dehumanisation is often achieved by 'describing them as either biologically subhuman ("cockroaches", "microbes", "parasites", "yellow ants"), mechanically inhuman ("logs", "packages", "enemy morale"), or supernaturally alien ("devils", "Satan", "demons")'—and has been used historically to represent a minority as an existential threat to the majority,

> Dehumanising discourses and conceptions have been identified in almost all major mass atrocities, prominently including those of Nazi Germany, Sta-

> linist Russia, Rwanda, Yugoslavia, Cambodia, Indonesia, and the Japanese occupation of China. Often, outgroup members (or victims-to-be) are even compared with toxins, microbes, or cancer, suggesting that they are polluting, despoiling, or debilitating the entire in-group—leading to particularly prominent recurring demands to 'purify' groups or societies from the supposedly toxifying elements (80).

'Dangerous speech', a category that has been expounded in detail by Maynard and Benesch (2016), is speech that constructs an 'outgroup' as an existential threat to the 'in-group', whether this threat is real or otherwise (81). Maynard and Benesch empirically identify the range of techniques commonly used in dangerous speech.

Dehumanisation and another technique called 'threat construction' are often inextricably linked, given that 'where dehumanization makes atrocities seem acceptable, threat construction takes the crucial next step of making them seem necessary' (82).

The field of resources that deconstruct and define the appearance of dehumanization online is still in its infancy.

Haslam's (2006) model that proposes links between conceptions of humanness and corresponding forms of dehumanization provided further detail for a theoretical base of this study's discourse analysis. Like Maynard and Benesch, he refers to 'animalistic' and 'mechanistic' forms of dehumanisation but details the characteristics that underpin both. If a subject is dehumanised as a mechanistic form, they are portrayed as 'lacking in emotionality, warmth, cognitive openness, individual agency, and, because [human nature] is essentialized, depth.' A subject that is dehumanised as animalistic, is portrayed as 'coarse, uncultured, lacking in self-control, and unintelligent' and 'immoral or amoral' (258).

There is also still limited understanding of the influence of dehumanising content on the specified target group, in this case, Muslims, and any contribution to cumulative or reciprocal extremism (Abbas 2020).

Online dehumanisation, disgust, and wanting to harm are also reflected in offline verbal abuse and threats, demonstrating how social media directly contributes to real world violence in Australia (Iner 2019, p. 9) and overseas (Muslim Advocates and GPAHE 2020).A vehicular terrorist attack on an entire Muslim family in Canada, a violent assault of a 38-week pregnant Muslim woman in Australia, the vandalism of an Australian mosque calling Tarrant a Saint and signalling to the Bosnian genocide and Christchurch attacks with "Remove Kebab", are all indicative of the way that Muslims have been divorced from the human family in the minds of violent individuals. Community submissions in Australia have reflected reports by young Muslim girls that the moment they begin wearing hijab, some members of the public cease to regard them as children, let alone human (Jabri-Markwell and Hashimi 2021).

A policy proposal from Australian researchers and practitioners to the Global Internet Forum to Counter Terrorism (Risius et al. 2021), suggested that serial or systematic dehumanization of an outgroup should be used as a definitory factor to distinguish violent extremist content from fringe discourse. There, dehumanization was characterised as an ideologically sanctioned form of 'non-physical violence'. Risius et al. write:

> The normative context of dehumanization establishes social preconditions within which violence by extremist instigators is likely to be perceived as justified. They authorize individuals to perform violence and shape bystanders' reactions to these events, and establish the parameters for depersonalization and stigma (Goffman 2009) or dehumanization and moral exclusion (Bandura 1999).

Building upon these understandings, it is hoped this research will encourage digital platforms and regulators to more effectively intervene in dehumanisation, both as a harm and known pre-cursor to atrocity. Whilst this study brings together and builds on genocide prevention, psychology and discourse analysis fields, it may be one of the first English-language studies to explore the function of dehumanisation within purposed online information operations and modern online discourse.

Methods

This research utilises qualitative discourse analysis to analyse the outputs of certain actors shared on Facebook and Twitter, as well as a mixture of discourse analysis and ethnographic content analysis to consider the written responses by social media users to those outputs. Ethnographic Content Analysis (ECA) allows for an evolving line of enquiry in reflex to observations about patterns of communication, meaning and behaviour. In this case, the categories for comments were initially shaped and defined, but through observing comment threads and their discursive context over time, we allowed for the categories to develop further to more precisely reflect patterns of meaning being communicated. Embedded in the constructivism–structuralism traditions, discourse analysis's key emphasis is on language in a social context.

Selection of actors

We used purposive sampling to identify information-rich cases related to the phenomenon of interest to select the actors.

This study focuses on one main a seed site identified in a detailed study undertaken by Benjamin Lee in 2015. Moreover, this actor was chosen as a main focus because it produces a significantly higher and more consistent volume of articles compared to the others. This seed site is the main production arm of Actor A, which is recognised as a hate organisation by researchers (ISD and GDI 2020, p. 26), but not Facebook and Twitter. The authors began by studying the comment threads of this Actor to determine if platform tools were able to identify explicit dehumanising language, as defined earlier in this paper. Noticing the ongoing prevalence of this language, the authors expanded the focus to consider other language signals Actor A may be using over time to dehumanise their targeted out-group. This led the authors to consider the ideology espoused on their home web site, as well as patterns of language in the headlines frequently shared. Following this route, the authors considered language techniques that might over time reinforce an underlying dehumanising discourse without triggering platform detection.

In order to illicit further insights, this study also examined the outputs of four (4) other actors to compare their respective use of traditional (explicit dehumanising language) and alternative (other language and discourse signals) methods of dehumanisation. Thus, in this paper,

- 'Actor A': David Horowitz Freedom Center is responsible for the project of 'Jihad Watch' website. It is US based.
- Actors B ('BareNakedIslam'), C ('Creeping Sharia') and E ('Richardson Post'). Actor E is Australian based. Actors B and C appear to be American.
- Actor D: The identity of this actor is unknown. Actor D's national location is not declared.

Sampling techniques for URLs

Headlines feature prominently in the tweet/posts, are an important signal of discourse (van Dijk 2008), and the gateway by which these seed sites draw social media users into the article content itself.

In the first sample, Actor A headline URLs ("URLs") were identified according to their prominence in Facebook pages found through a Victoria University study (Peucker et al. 2018) over a period of four months (March to July 2020). In subsequent samples, all URLs pertaining to Muslims and Islam published by these sites within defined periods, ranging from 5 days to a month, were collected. Actor C did not produce many URLs over a fortnight in Sample 5, so a second sample was taken of Actor C over a longer period (a month). Samples of Actor D and E were obtained through different methods outlined below.

For Actor D: Using Crowdtangle software, 193 referring public Facebook pages and groups of Actor A URLs from Sample 1 were identified. Ethnographic content analysis was conducted on these public pages and groups through reviewing the patterns of posting and communicated meaning, which helped to identify 31 pages and groups that primarily existed to portray Muslims as subhuman or existential threat to western civilisation. In October, a more detailed ethnographic analysis was conducted of the pages and groups, resulting in the identification of the series of related URLs from Actor D. These URLs were compiled into Sample 7.

Actor E URLs were identified by observing one particular Facebook page recommended to the researcher by Facebook whilst analysing the comment threads of top-performing posts from other Actor samples.

This combination of sampling methods was used to examine actors referred to in research and those who have not but appear on the face to exhibit similar behaviours. The sample sizes varied based on the outputs of Actors in those windows of time. See Table 1.

Method for seed site analysis

The home websites of actors (also known as 'seed sites') sometimes openly summarised their ideology, which we subjected to language and discursive analysis. The interpretative tool of personification was used to identify where Islam was being used as a proxy for Muslims. This is explained further detail in the Results section.

Method for headline analysis

The participant (nouns) and processes (verbs) of the headlines were extracted and analysed to understand the positioning of the subject over time, and to investigate 'who' the subject was within the Actor's discourse and its effect in constructing group identity.

The content of the headlines was also analysed for dehumanising descriptors, synonyms, and any coded language indicators pointing to Right-wing extremist (RWE) narratives. Connection to RWE narratives was determined based on narratives identified in extant literature (Lee 2015; Peucker et al. 2018; Davey and Ebner 2019).

Sampling technique for comment threads

Due to the sheer numbers of referrals and comments, Crowdtangle software was used to determine the top five 'performing' posts on Facebook, and tweets on Twitter according to the number of highest interactions. In the October sample of Actor A, in order to examine a range of comment threads from a broader range of actors, the authors included the highest performing thread of any referrer, who had referred more than three (3) times from this Actor.

Method for comments analysis

The comments on the top performing posts and tweets were then qualitatively analysed with regard to their discursive context, and categorised according to (1) 'dehumanising speech' which included clear dehumanising language and dehumanising conceptions of Muslims such as demographic invasion theories; (2) 'expressions wanting to expunge or deport Muslims from society'; and (3) 'expressions of wanting to kill or see Muslims dead'.

In addition, all comments that expressed a desire for Muslims to be killed, or that glorified the death or genocide of Muslims were reported to the respective platform, and any responses recorded.

Due to the scale of work involved in analysing comment threads, we decided not to include analysis of the *content* of all articles from the URLs included in this study. A detailed, but more manageable, qualitative analysis of a sample of Actor A's articles (published in the period of August 2020) was conducted instead, which

Fig. 1 From Facebook page that routinely disseminates actor E URLs

confirmed the site continues to propagate the themes identified in work of Benjamin Lee (AMAN 2020). However, it must be noted that fact-checking, or the degree to which the articles published disinformation, was not analysed. This could form the basis of future research.

Results

Explicit dehumanising language analysis—main themes

Theme 1: Detection in comment threads is poor

Explicit dehumanising language ('invaders', 'disease', 'savages') directed at Muslims is frequently not detected by Facebook's and Twitter's tools. In June 2019, an Australian Facebook page with more than 120,000 followers, known to routinely share Actor E articles, shared a poster entitled "The Great Replacement" (see Fig. 1).

The meme was accompanied by similar derogatory statements implying that Muslims plan to conquer countries like Australia through higher fertility rates. The intense reactions to this poster were revealed in the extensive comments, with a significantly high proportion employing *explicit dehumanising language* ('Islam is a cancer on global society for which there is no cure', 'You import the 3rd world you become the 3rd world. And when they become the majority then what next? They won't have whitey to leech off. Just like locusts, infest & strip everything until there is nothing left', 'Deport the PEDO crap', 'They breed like rats', 'muslums…. reminds me of aids'), *expressions of wanting to see Muslims killed or dead* ('Shoot the fuckers', 'Drown em at birth', 'Society should start culling the Muslims', 'I think I now understand why during the serbian/croat the serbs culled the women') *and*

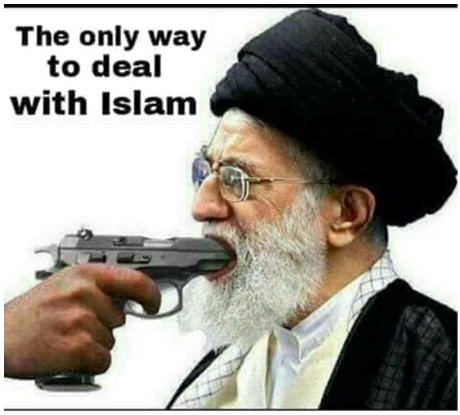

Fig. 2 From Twitter in response to an actor A URL

veiled threats of violence ('I'm going out tonight to do as much as i can to solve this problem'.)

Ethnographic analysis of another Australian page with more than 110,000 likes revealed the same pattern of page behaviour and user responses. This page relied on a steady flow of Actor D articles to generate fury, contempt and disgust towards Muslims amongst its audience.

This study found all of the Actors that attracted comments had a pattern of eliciting explicit dehumanising remarks, extremist ideology-based remarks concerning the target group as an existential threat, or glorification of, or incitement towards, violence against the target group. Actors A, D and E attracted long comment threads, especially in well-populated Facebook pages, groups and Twitter accounts. Actor C posts had comparatively fewer comments, and Actor B posts had very few comments.

As at mid 2021, the above post (Fig. 1) was still publicly available on Facebook, despite both pages being reported to Facebook's Public Policy Australian Manager for investigation in August 2020, alongside 30 other public pages and groups.

Given the scale of comments analysed in this study, many dehumanising attacks, but *all* expressions of wanting to kill or see Muslims dead, or glorifying the genocide of Muslims, were reported to Facebook and Twitter. Of all these reports, only a very marginal few were actually removed. This poor success rate underlines how automatic reviews struggle to accurately assess comments that are framed in response to material. Moreover, there were additional challenges to identifying dehumanisation in the anti-Muslim context.

Use of Islam as proxy to attack Muslims This study has found that terms such as 'Islam is a disease', 'Islam is a cancer', 'Kill Islam', and 'Exterminate Islam' are acceptable to both Facebook and Twitter based on testing of the platforms' reporting tools.

When Islam is called a disease or cancer, it is conceived in a way that is growing and spreading which happens on account of its number of followers growing. Thus, calling Islam a disease or cancer involves an attack on Muslims. This is supported

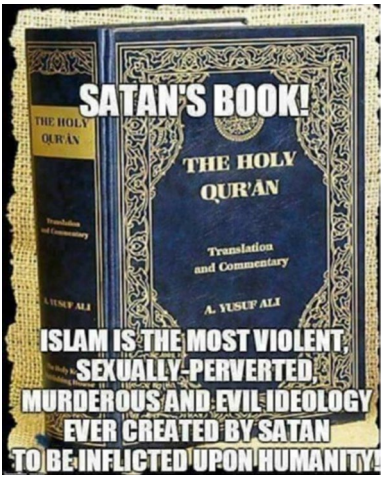

Fig. 3 From Facebook page in response to an actor A URL

Fig. 4 Muslims and Turks portrayed as dog and monkey like on Twitter

by the commentary about what must be done to 'treat this cancer'; which always involves restrictions, deportation, or annihilation of Muslims (Fig. 2).

Further, Islam is attributed human characteristics and actions, revealing this attack is directed at Muslims: 'violent', 'sexually perverted, murderous' (see Fig. 3).

Allowances for hate directed at criminals Facebook's hate speech policy protects members of certain groups based on protected characteristics like their religious affiliation, race, sexuality, but it does not extend to subsets of those groups, such as criminal elements. A user comment on the Paris beheading story of Actor A stated, 'Time to behead all paedophile moslems. NOW….'. Facebook found this to be consistent with its community standards, most likely because the term 'paedophile' was included before 'Moslems'. However, read contextually, Muslims are frequently referred to as paedophiles amongst these audiences, and the user in this case was responding to a story about a murderer, not a paedophile. These factors suggest the user is calling for all Muslims to be beheaded, not only a criminal subset, and this is how it would be readily interpreted by others on Facebook, especially by that audience.

Fig. 5 From Facebook page, in response to actor A URL

Memes and cartoons Automated systems often overlook images. For example, see Figs. 2 and 4 that Twitter did not respond to, even after they were reported. Figures 1 and 3 were reported to Facebook and remained online.

Theme 2: Materials that trigger explicit dehumanising responses tend to take advantage of hot political currents or serve ulterior political purposes

Some different sampling techniques were used in October to see which Facebook pages had shared at least three Actor A URLs in a sample period of 5 days. The largest reaching anti-Muslim pages included one page with 15 administrators across multiple continents (with 66, 000 followers), Actor A's own Facebook pages (amounting to more than 113, 000 followers), an Australian page (with more than 147, 000 followers), and a Canadian page (with over 35, 000 followers). The following headlines in that sample produced some of the highest interactions: 'Joe Biden vows that Muslims will serve "at every level" of his administration' reached a potential audience of 429, 572 followers (not including private groups or personal pages). It prompted a stream of fury from its audience (e.g. 'that sounds like aiding & abetting a TERRORIST denomination') and the proliferation of memes in the comments and pages: Fig. 5

Responses on Actor A's Twitter account to the same article referred to Muslims as a 'virus' and 'disease', including expressions that this decision would lead to violence (e.g. 'More crime, more Beheadings, more female mutilation, more honor

killings'), and that it was an example of 'stealth jihad'. One comment on Twitter also tagged US politician Ilhan Omar, stating that, 'Omar should be given an important position so she can promote her favourite practices like Sharia, polygamy and incest.'

Theme 3: There is a correlation between dehumanising language, extremist ideology and threats of violence

Audience responses to the article 'Paris update: Muslim beheaded teacher in street because he had shown Muhammad cartoons in class', shared Actor A's Twitter account, included dehumanising references to Muslims (separate to the murderer) as a cancer, virus, animals, and savages, and spawned significant commentary on the 'existential crisis' faced by France and the Western world from Islamic invasion, aided by liberals and the political establishment (with exception of Trump).

Audience responses to the same article on Facebook also revealed how these captured audiences interpret acts of terrorism and extremism conducted by ideologically motivated Muslims, and the frequent tendency to blame all Muslims and Islam, rather than the perpetrators alone. However, in this example on Facebook, it also escalated quickly to fantasies about violence. On Actor A's Facebook page, users responded with *dehumanising insults* ('They are worse than rabid animals, no brains of their own and vile to the core', 'MOSLEMS ARE INCOMPATIBLE WITH HUMANKIND', 'never trust them they are two faced. Like two people in one being. Ultimately their loyalty is towards Islam which is evil. If they never change their views on Islam no Matter how friendly, caring, compassionate they seem. If it came down to it they can become the most evil vile & depraved creature'); *calls to expunge Muslims* ('Do not let this atrocity happen in the US, vote the squad out, they are the enemy of mankind'); *repetition of demographic invasion/white genocide theory* ('They don't come to assimilate into western society, they come to dominate and conquer the infidels!! Wake up sheeple, these are barbarians!!', 'The ppl of Europe have to be detoxified from the twin evils of multiculturalism and diversity and then get rid of the leaders that spew lies and willingly put their own citizens to danger and evil'); *glorification of genocide of Muslims* ('The muslims are the only people on Earth who will earn their genocide, but they will be the only genocided people for whom nobody will have a drop of tear'); *calls to war* ('Europe has been Invaded and occupied by Muslims, who have claimed Europe as theirs, since they have Proclaimed Sharia Law! NATO will have to declare War on the European Islamic Caliphate and Attack European Muslim Strongholds, if they want to become an Independent Europe again?', 'this cult should have its head cut off before it is too late, have you ever thought about when the oil runs out this cult will be looking at us, and they will show no mercy'); and *calls to vigilante violence* 'Servicemen only ask: CAN WE GO KILL THESE FUCKERS YET …….. Barbarians/E.F.Whulfh' posted by a user along with the meme below, Fig. 6.

In one Australian Facebook page that routinely shares Actor A articles, the users responded to this article about the Paris beheading with: 'Go in hot an shoot the lot' (which attracted 7 'like' and 'love' reactions), 'U let them in, they multiply rapidly n impose their will on you. High time France takes the upper hand. Learn from China and Russia.'

A steady flow of these articles appeared to prime audiences to use explicit dehumanising language and become susceptible to incitement (Maynard and Benesch 2016, p. 79).

Theme 4: Article headlines don't use explicit dehumanising language to trigger such language in the comment threads

Given the dehumanising language in the comment threads, it was then investigated whether the headline of the stories that spurred these responses also contained explicit dehumanising language.

In the studied samples, Actors A, C, D and E avoided using explicit dehumanising language for Muslims in their headlines; it is suggested that this was to appear more reliable and objective. Additionally, it would reduce the likelihood that sharing these articles will trigger hate speech detection on social media. In contrast, Actor B relied more heavily on dehumanising slurs in the forms of synonyms and adjectives. For example, a variation of 'illegal alien Muslim invaders' was frequently used instead of 'Muslims'. It used the descriptor of 'frothing-at-the-mouth' to describe a Muslim carrying out violence. Interestingly, Actor B also conjured far fewer comments on Twitter. It is possible that the explicit dehumanising language in the headline made it unnecessary for the user to voice their disgust. It is also possible that there are other algorithmic factors at play, which may have influenced the number of comments.

Dehumanisation without the explicit dehumanising language—Main themes

Theme 1: Dehumanising conceptions can be present in ideology and work to dehumanise an out-group to an in-group audience.

Staying with this expanded 'roots and leaves' focus, we then considered whether there could be other signals or properties in the posted content that prompted responses to use explicit dehumanising language. Our first step was to return to their home websites.

Actor A is a central figure of the 'counter jihad' movement. Scholars like Benjamin Lee (2015, pp. 251–3), Melagrou-Hitchens and Brun (2013), and others, classify the 'counter-jihad' movement as an extreme right movement. Their site relies on a heavily skewed misrepresentation of Islamic theology to advance a demographic invasion narrative. Its predominant themes can be demonstrated in their 'FAQ' and 'Islam 101' pages. In this text, 'Islam' is personified by the attribution of human actions and qualities to the religion as a whole [emphasis added]:

- 'Islam *exists in a fundamental and permanent state of war with non-Islamic civilizations, cultures, and individuals* (a group of people, not a religion, can be in a state of war with civilisation)',
- 'A halt to terrorism would simply mean a change in *Islam's tactics*—perhaps indicating *a longer-term approach that would allow Muslim immigration and*

higher birth rates to bring *Islam closer to victory* before the next round of violence',
- 'Islam proper *remains permanently hostile'*, and
- 'Islam's *violent nature* must be accepted as given'.

This personification of Islam in this discourse enables Muslims to be portrayed as an existential threat implicitly, without falling foul of hate speech, as on the surface, a religion—not people—is the subject of attack. This distinction between content and discourse analysis that this paper seeks to highlight.

Dehumanising conceptions in this studied discourse include the portrayal of Muslims as:

(1) 'mechanically inhuman' (Maynard and Benesch 2016, p. 80)'theological automatons' who are 'unified in thought and deed' to carry out demographic invasion (Lee 2015, p. 252). Significantly, it follows that there is no way to tell if Muslims are truly peaceable or not.
(2) 'Subhuman' (Maynard and Benesch 2016, p. 80) in their inherent violence, barbarism, savagery, or plan to infiltrate, flood, reproduce and replace (like disease, vermin).

Further, the site engenders a perception of legitimacy by seemingly engaging with primary texts of Islamic jurisprudence. However, extreme right-wing actors often authored the material presented rather than stemming from genuine engagement with academic scholarship. For example, of great concern was the publication of Bat Yeor's work in the 'Frequently Asked Questions' and 'Islam 101' pages of the site. Yeor is the original author of the Eurabian conspiracy theory, whose ideas were heavily drawn upon by Norwegian terrorist Anders Breivik (Archer 2013; Berntzen and Sandberg 2014). In this regard, social media platforms could also take legitimate action against Actor A based on *disinformation*, as studies have shown that personal religiosity and spirituality in Islam is inversely related to violent extremism (Beller and Kröger 2018; Aly and Striegher2012) and that terrorism by ideologically motivated Muslims overwhelmingly targets Muslim victims (and does not substantiate 'a clash of civilisations') (Cordesman 2017).

Comments made directly on Actor A's website were also analysed to provide further insight into the community and the effect of these articles on its readers in relation the October sample of Actor A. For example, dehumanising insults focused on Muslim men as sexual deviants ('Muslim men do not have sexual relationships with women. They rape them. That is all they know how to do'); and Muslims as subhuman ('Muslims are the dregs of the world and 90% don't even qualify as humans', 'mad dogs', 'animals', 'brutal Muslim beasts'). Further, many users employed the dangerous speech technique of threat construction to make violence and war against Muslims appear legitimate, proportionate, and in some cases, even necessary. Reader comments frequently reflected the premise of the site, that religiosity in Islam leads to sub-humanity and extremism.

Ideological assessment of Actors A and B was most straightforward, as both sites openly summarised their rationale. Actor C's site conveyed their rationale indirectly: For example, a list of 'Muslim enclaves in America' and menu choices pointed to Muslims as the enemy within or outside.

For Actor E, an Australian website, ideology was assessed by the prominence of demographic invasion and cultural genocide/suicide arguments in published opinion pieces, including by the site's editor. Examples include statements like the following,

> 'Your women will be taken, raped, sold as slaves and forced to breed new soldiers for Jihad, because he did it—for years. ... Next for the men, you will be beheaded or crucified if you don't join the jihad and agree to butcher your neighbours—for the glory of "Allah".'

> 'Turkey provides a particularly gruesome example in the late 19th and early 20th century for what happens next to Jacinda's daughters and friends when Muslims rise up in New Zealand. ... Christian women were crucified naked, the Muslim way, where a sharpened stake was inserted into their vagina and hammered up through their abdomen for a slow, painful death.'

> 'These 50 migrants flown in from Greece were a hand-picked group featuring women, a few children and younger-looking lads. ... The ones that follow will not look like this. They will be fighting-age men with the cunning, guile and aggression it takes to knife and claw their way to the front of a violent queue.'

When repeated throughout Actor E publications, statements like these create an increasingly irreversible picture of all Muslims as inherently barbaric.

A network of related sites was labelled Actor D. This collection of 'news' sites appeared to have the same creator and carried titles related to 'free speech', 'the real news', and 'the truth'. However, almost all stories lacked an identifiable author and the sites themselves did not identify an editorial team. The headlines relating to Muslims and Islam shared much in common with other actors in this study. Notably, links to this collection of websites were blocked on Twitter but remained widely shared on Facebook. Actor D articles were posted daily to dedicated anti-Muslim and far right Facebook communities.

From this exercise, it was clear that only some websites openly surmised their ideological position, whereas for others, it had to interpreted through qualitative assessment of their work. Moving forward, there needed to be universal markers that could measure dehumanisation as part of that qualitative assessment.

Theme 2: Headline language can dehumanise an outgroup through choice of participants and verbs over time.

The following step was to look at the features of the outputs of Actor A, to consider whether there were other language signals that constructed the target group as subhuman or inhuman. We began where most sentences begin, with the participant (also known as the subject).

Essentialising 'out-group' identity For Actor A, which purports to be concerned with Islamic demographic invasion and violent jihad, most of its headline subjects involve a wide net of Muslim identities. Whether a Muslim subject is serving in an administration, a progressive politician living in the West, an ordinary person seeking to migrate to another country, committing a heinous crime, expressing controversial or offensive views, or seeking human rights protection; whether they live in the US, Europe, India, Australia or the Middle East; or whether it is a Muslim majority nation state like Turkey, Pakistan or Iran, they're identified and centred as a the 'Muslim subject'. Subjects used in Actor A's headlines included:

> 1st sample: Muslim, Islamic movement, Muslim Cleric, Intifada, Islamic state, Iran-backed Jihadis, Iranian jihadists, Health Jihad, Muslim preacher, Muslims, Muslim students, Muslim leader, Muslim software engineer
>
> 2nd sample: Muslim, Muslim cleric, Niqab-clad Muslima, Muslima, Hamas-linked CAIR, Muslim migrant, 'guardian of morals', Illegal Muslim boat migrants, Egyptian migrant, Muslim professor, Jihadi murderers
>
> 3rd sample: Muslim, 'real victims' according to CAIR, Islamic Terror Sheikh, Muslim population, Muslim leader, Muslim migrant, Islamic State, migrants, Muslim migrant camp, Muslim cleric, Muslim lawyer, Shi'ite Muslims, Muslim, Muslims
>
> 4th sample: Muslim migrant, Muslim terrorist, Convert to Islam, Rashida Tlaib and Ilhan Omar, Muslim voter, Imam, Muslim Legal Fund of America, Wife of Orlando jihad watch mass murderer, Large Muslim population, Father of Muslim, Grandparents, brothers of Muslim, Muslim family, Black Muslim bakery, Muslim mass gatherings, Muslim judges, Muslim Arab youths, Illegal migrants, Islam, Mass migration, Muslims, Muslim cleric, Muslima

The melding of Muslim identities into one harmful threat is necessary for the argument that all Muslims are potential terrorists or terrorist sympathisers; and any Muslim, on account of their faith, could be inherently disloyal, deceptive, and harmful to their home country. This conception is built on the inaccurate premise that Islam itself is what radicalises people towards violence. This form of dehumanisation is achieved *in part* through the essentialising of Muslim identity, with cumulative headlines as a vehicle to its dissemination and acceptance within wider society.

Similarly, Actors B, C, D, E seemed to draw on a wide net to essentialise Muslims as a hostile mass to their in-group audience, so this was a clear indicator of such purposed information campaigns.

Negligible focus on ISIL/Al-Qaeda Overall, in the samples studied pertaining to Actor A, it was interesting to note that there was an extremely limited focus on ISIL and Al-Qaeda in the headlines. The two times ISIL was referred to by Actor A, was to propagate its message or connect it with Islam (i.e. 'Islamic state hails 9/11 as 'pivotal moment for Islam', and 'France: Man converts to Islam, becomes torturer and executioner for the Islamic State'). Actor A shows a strong preference for referring to

Muslims as the participant in its headlines; for example, 'Cameroon: Muslims target Catholic mission, murder 28, including 7 children ages 3 to 18'. ISIL and Al-Qaeda were even more absent in Actor B and Actor E headlines, however Actor C relied on references to ISIS more (8 out of 60 headlines).

The cumulative impact of dehumanising verbs in headlines Following our focus on the participant, we examined patterns in the verbs. Narrative writers know that carefully selected verbs and actions are a powerful tool to vividly portray a character—more so than any number of adjectives. Yet platforms are focused on dehumanising slurs that come in the form of synonyms and adjectives, as opposed to verbs. On average, Actor A produced 6–7 articles a day. The following verbs were attributed to a 'Muslim' participant in their headlines:

1st sample: Threaten to kill, sexually assaults, attack, murders, complaint, kidnap another, join forces, forcibly convert, kidnapped women and children held, breaks into, rapes, blames, torturer and executioner, spread the virus, 'punishes' you, ransack government buildings, threaten American facilities.

2nd sample: Plant IED, six children killed, Smashes Hindu idols, cheer scream Allahu Akbar, Sues McDonalds franchise for discrimination, Screams Allahu Akbar, struggles violently with police, committed the most extensive spree of felonies by a congressperson in US history (Ilhan Omar), threaten 'too Western' women, hang their pictures in mosques, Kidnaps minor Hindu girl at gunpoint, target Catholic mission, murder 28 including children ages 3 to 18, holds hostage at knifepoint, plotted to murder, revolt aboard Italian coronavirus quarantine ship, abandons LGBT alliance, feign reform, fools the establishment.

3rd sample: Screaming Allahu Akbar, Stabs, critically injures, Stabs, Escape, Attacks, sets fire, says it is a 'hate crime', Collects $ 35 000, Migrate, File complaint, made jihad bomb threat.

4th sample: Murders priest, Beheaded, Beats woman to death, Turned to terror, Call for high Muslim voter turnout, Screaming Allahu Akkbar beheads man, threatens police, is 'suspected terrorist', Beheads, Beheaded, Killed, Gets 28 years for hammer attack, Arrested , Will serve at 'every level' of his administration, Beats woman to death, Stabbed, proud of son, Beheaded, Arrives from Iraq in private jet, Proud of trying to honor-murder ex-wife, Tried to steal $ 22 million, but isn't killing people anymore, Go on, Want to migrate to the west, Destroyed Europe, Open fire on bus, try to separate passengers by religion, say they won't administer marriage ceremonies if celebrations include music and dance, ransacks supermarket, allowed to keep veil.

Actor A relied on negative verbs to associate Muslim identity with sub-humanity, sexual deviancy, and construct them as an existential threat. This pattern was found across the Actors, with exception of Actor C, who tended to use more legal and technical verbs, as well as passive voice, which muted the impact of their verbs ('arrested', 'indicted', 'testifying', 'charged', 'linked').

Actor B's reliance on dehumanising verbs to convey sub-humanity over time was as prolific as Actor A—for example: 'Doesn't hide his disgust for non-Muslims', 'Set fires', 'Beat up a white guy', 'Smuggling', 'Threaten', 'Burn down', 'Put up in 5 star luxury hotels', 'Demands 'hate crime' investigation', 'Brutally stabbed', 'Killing', 'Critically injuring', 'Replaced them'.

Actor D attributed the following actions to a range of Muslim actors: 'Explains that child marriage is perfectly fine', 'Complain about life in the UK', 'Murdered his daughters', 'Abducted and raped 6-year-old Christian girl', 'Calls to burn women's face with acid if they refuse to wear hijab', 'Promises terror attacks', 'Killing Jews', 'Starves his 13-year-old wife', 'Murdered in acid attack', 'Celebrates as he marries 14-year-old Australian girl'.

Actor E tethered Muslim actors together with actions such as 'hacked to death and beheaded' and 'brutally murdered'.

The cumulative use of dehumanising verbs was strong indicator of a purposed information campaign.

Theme 3: Coded language acts as a touchstone or lightning rod for dehumanising conceptions

To a significant degree, Actor B's headlines, and to a smaller degree Actor A and Actor E's headlines, used coded language invoking extreme right demographic replacement theories. Such language included 'no go zone', 'invader', 'invasion', 'sharia Sweden', 'anti-Islamization', 'cultural enrichment', 'colonisation', and 'conquest'. Actor B referred to Muslims as 'invaders' in a third of its sampled headlines. Christchurch terrorist, Brenton Tarrant, also referred to Muslims as 'invaders' in his so-called manifesto. The use of coded language serves to make such terms mainstream, but also acts as a lightning rod for user comments expounding upon these theories.

Theme 4: 'Accusation in the mirror' and dehumanisation work hand in hand.

Characterising a group as mechanically inhuman (and thus incapable of independent thought or feeling) is dehumanising. But it's also foundational to a technique called 'accusation in the mirror', used to incite violence. Maynard and Benesch explain 'accusation in the mirror' as a technique designed to construct an out-group as an existential threat to the in-group,

> In a strange yet common form of threat construction, a speaker accuses another group of planning to engage in the sort of violence that the speaker wants to see perpetrated against them, instead. This has been dubbed (originally in a Hutu propaganda manual discovered after the 1994 genocide) "accusation in a mirror." [citations omitted] Examples of the technique are legion. In the speech with which we opened this article, Arthur Seyss-Inquart accused Jews of planning to annihilate the German people—a baseless claim that in fact mirrored what the Nazis planned and would attempt to do to the Jews. The idea that Jews would wipe out the German *volk*—if Germans did not pre-

> empt that effort—was a relentless feature of Nazi propaganda, of which Seyss-Inquart's speech was just a typical example. Speeches and articles by Hutu leaders before that genocide similarly warned that the Tutsi were planning to annihilate the Hutu (82).

The claims that Muslims are acting as trojan horse in the West, with the plan to wipe out all non-Muslims, has been used to justify violence by Breivik, Tarrant and others. Very occasionally, Actor A used phrases that could be characterised as 'accusation in the mirror'. For example, 'Erdogan threatens to declare a *religious war* against Christians after Austria closed mosques', 'Islamic Movement and Left Join Forces in *anti-American Revolution* (a 3 part story)', 'Australia: Muslim leader says Israelis "are waiting for the *Islamic nation to carry out the jihad against them*"'.

In addition, Actor B explicit labels Muslims as 'illegal alien Muslim invaders', as well as propounding invasion notions in headlines such as, 'UK: The *Muslim invasion* is real and exploding'. Actor E referred to '*Islamic colonisation'* and '*conquest*'. 'Accusation in the mirror' was not observed in Actor C headlines.

Theme 5: Many headlines triggered explicit dehumanising responses with value-neutral language.

Additionally, this study identified a further category of headlines that did not use explicit dehumanising language, whether in the form of verbs or descriptors, at all.

Most notably, in the September sample of Actor A, a majority of headlines did not use dehumanising descriptors, verbs or coded language (21 of 34 published posts); for example, 'Muslim leader says "Islam is the second religion in France. Those who do not like us have only to leave France"'. It did however generate some of the largest numbers of user comments on Twitter for the same sample, including explicit dehumanising language ('Stop it before this became forth stage cancer','blood sucking cult', 'Average muslim has genetic disorders due to consanginuity', 'who would make friendship with snakes?'), and iteration of demographic invasion theories ('in the name of secularism and liberalism u gave them shelter and in return they will loot u kill u convert u', 'The disease of Islam is poisoning Western society'). This headline also prompted ominous calls to expunge Muslims from France:

> 'Sad to see the demise of a great European country. This virus will kill every European nation in the next 30 to 50 years if not treated in time.'

> 'The French tried to be nice and show how accommodating they were, now they are about to reap the fruits. You still have time to spray "malathion" [an insecticide] and get rid of the weeds.'

> 'China, Israel and Myanmar are best examples of dealing with radical Islamic terrorism and Jihadist.'

> '@EU_Commission r u feeling bruh? where're u going to go on boats full of yourself becoming refugees, an outcome of letting these invaders? @francediplo_EN @French_Gov its not too late to buckle up, hustle & set law & order right. screw human right idiots'

> 'When will the world see that they multiply like rats. they will not let our future generations live in peace. The only solution is to separate them from our lands, just like what Poland does.'

Some Twitter users were more pointed in their calls for violence, which was heavily accompanied by dehumanising language:

> 'But this breed is neither going to change nor they can be restricted. There's only one solution to this..... You know what'
>
> In response to the above tweet:
>
> 'You know what, they will change but they need a treatment like Uighurs'
>
> 'The only issue is that normal human being can't go to that extent, not in our imaginations, that is the perk they are thriving on. When you become ruthless, make them untouchable, they will come to their sense'
>
> 'True. But this breed itself thinks that they are untouchables. They can't come to their senses because they are born non-sense.'
>
> 'Time for another round of crusades Really we need it Deploy Army and all fight against it'

In contrast, the Facebook user comment thread on this article had less dehumanising insults and more reiteration of the demographic invasion theory, lamenting the 'loss' of France and Europe, saying they [Europeans/Westerners] brought it upon themselves, and warning fellow Americans about the 'Muslim plan'. This one article reached an estimated audience of 366, 844 Facebook followers (not including private groups or personal pages).

The degree to which each actor in each sample relied on dehumanising descriptors, verbs, and coded right-wing extremist language varied significantly across the samples: Fig. 7

In the first Actor A sample containing articles primarily from July, these particular headlines leveraged anxiety about the recent Coronavirus pandemic to portray Muslims as an existential threat. For example: 'Pakistan: Muslims crowd into mosques, "We don't believe in coronavirus, we believe in Allah."', 'India Muslim software engineer: "Let's join hands, go out and sneeze with open mouth in public. Spread the virus"', and 'Italy: Muslim says coronavirus is "something from Allah, a positive thing" because "people are going mad"'. Here, the headlines allegedly quote individual Muslims to propagate the dehumanising discourse that Muslims act as a collective, without the capacity for independent thinking reflexes or human empathy.

Actor C focused substantially on charges, arrests, and anniversaries in relation to terrorism perpetrated by Muslims ('U.S. Has Repatriated 27 Americans from Syria and Iraq incl 10 Charged with Terrorism-Related Offenses for Their Support to ISIS Terrorists'), and linked this behaviour to Muslim identity and Islam more broadly through its implied concern about Muslims seeking election in the USA ('Terror-linked CAIR attempting to "Train 200 Muslim Candidates to Run for

Office"', 'Joe Biden says Muslims will serve "at every level" in his administration', 'Michigan: Gun-banning, Squad-supporting Muslim seeks House seat in District 4 (Hamtramck)'). Its reliance on explicit language to dehumanise Muslims was comparatively very low. Unlike the other Actors, Actor C was mainly focused on Muslims in the US. From its headlines, one could deduce that the greatest threat posed by Muslims to the US was not demographic invasion, but terrorism and 'infiltration' through the electoral system.

Without using explicit dehumanising language, Actor D successfully conveyed dehumanising discourse such as Muslims are barbaric ('Watch: immigrant in the UK explains that child marriage is perfectly fine' and 'Watch: 50-year-old Muslim Man starves his 13-year-old wife for disobeying him'), with demographic invasion/replacement narratives ('German mother cries "we feel like foreigners" as 2 of 25 in son's daycare speak German', and 'Sharia is gaining popularity in France as Jews are driven out').

The key insight from this analysis was that the marshalling of stories to create an overwhelming sense of crisis and disgust does not always rely on explicit dehumanising descriptors, verbs, or coded language in the headlines. It appeared that where an audience had been primed over time, implied properties in text were capable of triggering entire sub-texts.

Establishing proofs for extreme right narratives whilst maintaining value neutral language. From here, we hypothesised that actors employed apparently neutral headlines to act as a dog whistle or touchstone to underlying theories. Often these kinds of headlines prompted users to explain the dehumanising theory about the outgroup in the comment thread—either to helpfully connect the dots for other readers, or to demonstrate superior insight. All headlines were analysed for whether they offered 'proof' to extreme right narratives about Muslims and Islam Fig. 8.

The narrative that Muslims are seriously violent, barbaric, and subhuman was dominant for all Actors with exception of Actor B. Demographic invasion was the most significant narrative focus for Actor B, and equally significant focus for Actor E, and a consistently substantial narrative for Actor A. Proportionally, all of the Actors provided 'proofs' through their headlines for a broad range of extreme right narratives.

Baiting the 'in-group' audience could be done with value-neutral language. It was also clear that in-group audiences were being baited with headlines. For example, 'New book claims landmarks of Western church architecture were "stolen" from the Islamic World' was a form of 'baiting' an in-group, given their attachment to the idea that Western civilisation is superior, separable, and 'pure' of Islamic influence. Headlines that tap into the narrative that authorities are impotent or complicit in this threat create an overwhelming sense of rage for some members of the in-group. In the September sample alone, Actor A framed authorities in the US, UK, France, Sweden, India, and Denmark in this manner ('Cops insist it is a love affair and refuse to investigate', 'French cops in Muslim no-go zone: "You only have to move somewhere

else" or go vigilante', 'Fearing charges of Islamophobia government allows sharia marriages').

Treating user responses as a guide, a headline that tapped into in-group outrage that the outgroup was, in their distorted worldview, taking advantage of western liberal democratic largesse to culturally 'contaminate' and take over, was a powerful trigger. For example, Actor A's article 'Muslims migrate to Australia, file complaint with Human Rights Commission because food they're given isn't halal' produced numerous responses expounding on demographic invasion and white genocide. Common dehumanising conceptions from those on Twitter were that Muslims originate from 'cesspools', 'toilet bowl countries', and 'shitholes', and that resisting their plot had to be done for the sake of 'civilised world and culture'. It appeared to 'trigger' users who saw this as an attempt to 'placate the Moslem invaders'. One user commented, 'Physical appearance of mooslems is like normal human being but mentally like cold blooded demon, Ogre.' The world 'infiltrate' was preferred to migrate. Many spoke about the 'stages' of 'jihad' in taking over a country: 'It starts with halal food, next is burning cities and killing infidels.' Whilst others lamented that the west was contributing to its defeat: 'A secularism & multiculturalism is a breeding ground for deadly peaceful community virus (Islam).' The disgust prompted by this headline also led to calls to expunge: 'What are the options available with Australia? Will they let the cancer spread there also, like it has in Europe?'.

Actor A's baiting headline, 'Germany: Thermal spa in area with large Muslim population bans bikinis that are too tight' was one of the top performing articles on Facebook. This headline, without any dehumanising descriptors, verbs, or coded language, prompted responses from Actor A's Facebook audience that included dehumanising insults ('cockroaches', 'insiduous mould', 'trojan horse', 'piss on them', 'plague of humankind', 'quranimal'), calls to expunge and restrict Muslim immigration ('ban the satanic cult instead', 'fuck off back to a Muslim country you fucktards' and 'deport the pedophiles along with Merkel'), and calls to violence ('time for ethnic cleansing!').

Manufactured irony was a powerful baiting technique Quotes from Muslims that mirror how the in-group audience feels towards Muslims were favoured in headlines and produced strong user reactions. For example, the Actor A headline, 'UK Muslim preacher is Extremely offended by British Women Who Show Hair, Put on Makeup & Perfume in Public' triggered its audience with the terms 'extremely offended', prompting Facebook user responses with explicit dehumanisation language such as,

> 'I am deeply offended at the manner in which these moronic followers of the satanic cult founded by the peed0file false prophet Mu Ham Mad, are so easily offended in countries which they have deliberately invaded!'

> 'THEY OFFEND ME. THEY RAPE LITTLE BOYS AND GIRLS. THEY DO NOT RESPECT WOMAN. THEY HOLD THE WOMEN DOWN. AND THEY WANT EVERY INFIDEL [AMERICAN] DEAD! THEY SHOULD GO BACK TO THEIR MUSLIM COUNTRIES.'

> 'I'm offended by Islam and everything it represents , sick to death of the demands these Neanderthal inbreds make daily !'
>
> 'What DOESN'T offend these subhuman monsters???'
>
> 'Well I'm offered by Muslims. Their nothing but pig's.... No wait a min pigs are a lot better than Muslims. At least pig's don't f..k their young like Muslim men do'

This headline also prompted expressions of wanting to see Muslims dead: 'please just go put a C4 vest on and press the button', and 'NO, EXTERMINATE ALL ISLAM'.

Similar techniques and responses were observed on Twitter. A tweet which included a link to article titled: 'Turks irked at Qur'an-burning in Sweden: "We expect Swedish authorities to take all measures to **fight this disease**" [emphasis added] prompted numerous responses that Islam, jihad, and Turkish people were the disease, as well as other dehumanising references to 'refugee parasites', and Muslims as 'snakes'.

A tweet which included a link to an article titled: 'Sweden: Rioting Muslima screams "Why did they bring us here, the *stray dogs*, if they do not want Islam?"' [emphasis added] prompted responses characterising Muslims as dogs, snakes and rubbish. This tweet also generated comments inciting violence. One person commented with a meme showing a gun being placed in a Muslim man's mouth, with the words, 'The only way to deal with Islam'. Another user commented, 'Give them red hot copper capsules, they love it, it's the ticket to jannat for them.... They are the predecessors of desert cannibals', again showing the relationship between dehumanising insults and calls to violence. Twitter users also responded, 'Send them back to their own country or put them under detention till they die', and 'Dump them in the sea.'

These headlines manufactured a sense of irony that Muslims would be daring to accuse others of things that are 'true' about themselves, according to in-group members. This irony seemed to powerfully trigger those in-group audiences, stirring explicit dehumanising language and threats of violence.

Implications of dehumanising language analysis for understanding content on Facebook and Twitter

The presence of dehumanising *language* is unnecessary to propagate dehumanising *discourse*, yet it is the specificity of dehumanising language that social media companies largely rely upon to detect online hate actors.

This study underlines the need to adopt a different framework capable of assessing manipulative and dehumanising 'identity builds' of out-groups that go beyond explicit dehumanising synonyms or adjectives. As part of this framework, platforms might assess an Actor's information campaign by considering the following elements.

- In some cases, the ideology stated on actor websites
- whether the actor has personified and dehumanised a noun as a direct proxy for a group
- participants (nouns) and verbs of headlines that work cumulatively to essentialise and dehumanise the subject
- headlines that act as proofs for dehumanising conceptions and theories,
- headlines that bait in-group audiences
- and in some cases, a pattern of hate speech and other violent speech in the comment threads.

Whilst some information campaigns will have overt ideological agendas, and some won't, the above markers have the potential to act as cues for cumulative dehumanisation.

Counter-jihad ideology and dehumanisation

Moreover, this study raises questions about platform treatment of dehumanising *conceptions* about Muslims. It would appear that Facebook and Twitter are still unclear on whether 'counter jihad' ideology is harmful disinformation that dehumanises Muslims.

As discussed in this paper, Islam is frequently used as a proxy for Muslims. In counter jihad discourse, Islam is attributed human actions and qualities and then dehumanised, as a seemingly more liberal route to dehumanising Muslims.

Conclusion

This study relied predominantly upon qualitative analysis of the online content; however, such analysis could be upscaled in the future through natural language processing and machine learning, like the Institute for Strategic Dialogue has already begun doing with Twitter data (Davey et al. 2020).

Platforms' self-evaluation reports and mandated transparency reports by nation states may carry little value whilst systems do not detect dehumanisation. Furthermore, social media companies often lack explicable policies on how they assess 'hate' or 'dangerous' organisations to bar them from using their platforms. Given its established links to atrocities and genocide, dehumanisation offers a widely accepted measure. However, as this study suggests, its operation through discourse (not language alone) must be analysed by platforms to competently assess actors and their information operations (Figs. 5, 6, 7, 8; Table 1).

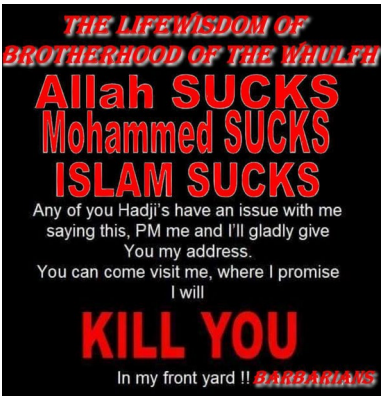

Fig. 6 From Facebook message posted in response to an Actor A URL

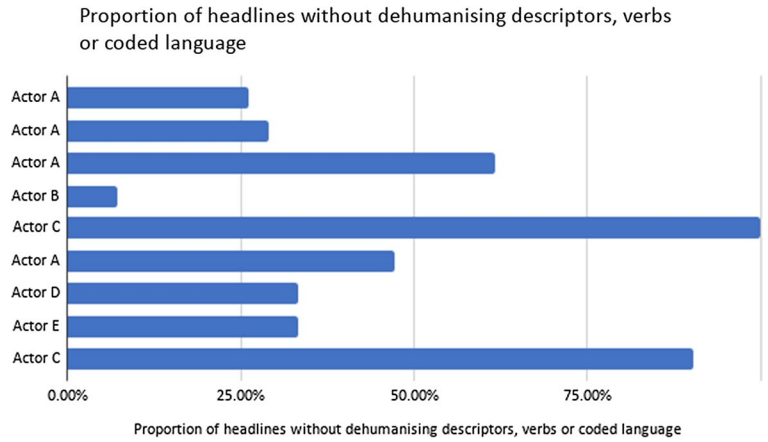

Fig. 7 Proportion of headlines without dehumanising descriptors, verbs or coded language

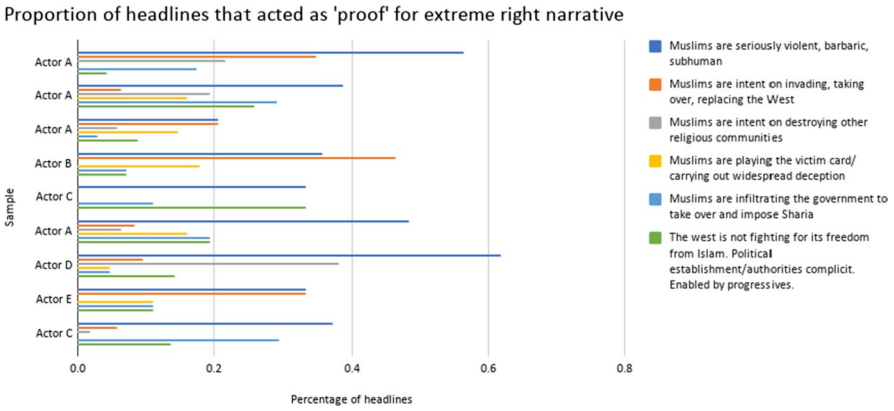

Fig. 8 Proportion of headlines that acted as 'proof' for extreme right narrative

Table 1 Outline of sample timing, size, source, and treatment used as a basis for this study

Sample (in the order taken)	Period sample covers	Number of URLs	URL source websites	Platforms their reach and interactions were tested for
1	29 March–13 July 2020	25	Actor A	Facebook
2	13–17 August 2020	31	Actor A	Facebook and Twitter
3	9–13 September 2020	34	Actor A	Facebook and Twitter
4	8–15 September 2020	28	Actor B	Facebook and Twitter
5	2–15 September 2020	9	Actor C	Facebook and Twitter
6	5–13 October 2020	36	Actor A	Facebook and Twitter
7	September and October 2020	21	Actor D	Facebook and Twitter
8	October 2020 (whole month)	51	Actor C	
9	March–August 2020	9	Actor E	

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Abbas T (2020) Far right and islamist radicalisation in an age of austerity: a review of sociological trends and implications for policy. International Centre for Counter Terrorism, The Hague

Allchorn W, Dafnos A (2020) Far right mobilisations in Great Britain: 2009–2019. Center for the Analysis of the Radical Right, London

Aly A, Striegher JL (2012) Examining the role of religion in radicalization to violent islamist extremism. Stud Confl Terror 35:12

Archer T (2013) Breivik's mindset: the counterjihad and the new transatlantic anti-muslim right. In: Taylor M, Holbrook D, Currie PM (eds) Extreme right wing political violence and terrorism. A&C Black, London

Australian Muslim Advocacy Network (AMAN) (2020) Interim Report: an introduction to extreme right actors and ideologies targeting the Islamic community

Bandura A (1999) Moral disengagement in the perpetration of inhumanities. Personal Soc Psychol Rev 3(3):193–209

Beller J, Kröger C (2018) Religiosity, religious fundamentalism, and perceived threat as predictors of Muslim support for extremist violence. Psychol Relig Spirit 10(4):345

Berntzen LE, Sandberg S (2014) The collective nature of lone wolf terrorism: Anders Behring Breivik and the anti-Islamic social movement. Terror Polit Viol 26(5):759

Cordesman A (2017) Islam and the patterns in terrorism and violent extremism. Center for Strategic & International Studies, London

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

Davey J, Hart M, Guerin C (2020) An online environmental scan of right- Wing Extremism in Canada: an Interim Report. Institute for Strategic Dialogue

Department of Security Studies and Criminology (2020, October 9) Mapping networks and narratives of online right-wing extremists in New South Wales (Version 1.0.1). Sydney: Macquarie University

Facebook (2020) Community standards

Goffman E (2009) Stigma: notes on the management of spoiled identity. Simon and Schuster, New York

Haslam N (2006) Dehumanization: an integrative review. Personal Soc Psychol Rev 10:257

Iner D (ed) (2019) Islamophobia in Australia Report II 2017–2018. Charles Sturt University and ISRA, Sydney

Institute for Strategic Dialogue (2020) Lens on hate (weekly series in lead up to 2020 US Election)

Institute for Strategic Dialogue and Global Disinformation Index (2020) Bankrolling bigotry: an overview of the online funding strategies of American hate groups
Jabri-Markwell R, Hashimi M (2021) Inquiry into serious vilification and hate crimes (Submission to the Legal Affairs and Safety Committee of the Queensland Legislative Assembly, Australia)
Lee B (2015) A day in the 'swamp': understanding discourse in the online counter-jihad Nebula. Democr Secur 11(3):248
Lentini P (2019) The Australian far-right: an international comparison of fringe and conventional politics. In: Peucker M, Debra S (eds) The far-right in contemporary Australia. Springer, Singapore
Mac R, Silverman C (5 November 2020) Facebook has a metric for "violence and incitement trends". It's rising. Buzzfeed News
Marczak N (2018) A century apart: the genocidal enslavement of Armenian and Yazidi women. In: Connellan MM, Fröhlich C (eds) A gendered lens for genocide prevention. Palgrave Macmillan, London, pp 133–162
Maynard JL, Benesch S (2016) Dangerous speech and dangerous ideology: an integrated model for monitoring and prevention. Genocide Stud Prev 9(3):70
Meleagrou-Hitchens A, Brun H (2013) A Neo-nationalist network: the English Defence League and Europe's Counter-jihad Movement. London
Murphy L (2020) Facebook's civil rights audit—Final Report
Muslim Advocates and the Global Project Against Hate and Extremism (GPAHE) (2020) Complicit: The human cost for Facebook's disregard for human life
Peucker M (20 Oct 2020) Should we stop referring to some extremists as right-wing?. ABC Religion and Ethics
Peucker M, Smith D, Iqbal M (2018) Mapping networks and narratives of far-right movements in Victoria: Project Report. Institute for Sustainable Industries and Liveable Cities, Victoria University
Risius M, Blasiak K, Wibisino S, Jabri-Markwell R, Louis W (2021) Dynamic Matrix of Extremisms and Terrorism (DMET): a continuum approach towards identifying different degrees of extremisms. Report to the Global Internet Forum to Counter Terrorism
Twitter (2020) Hateful conduct
van Dijk TA (2008) Discourse and racism. In: Goldberg DT, Solomos J (eds) A companion to racial and ethnic studies. Blackwell, Malden
Wodak R (2015) The politics of fear: what right-wing populist discourses mean. Sage, London

Authors and Affiliations

Mohamad Abdalla[1,2] · Mustafa Ally[3] · Rita Jabri-Markwell[4]

Mustafa Ally
Mustafa.Ally@usq.edu.au

[1] Centre for Islamic Thought and Education, UNISA, Adelaide, Australia

[2] School of Education, Victoria University of Wellington, Wellington, New Zealand

[3] Information Systems, School of Management and Enterprise, University of Southern Queensland, Toowoomba, Australia

[4] Australian Muslim Advocacy Network (AMAN), Brisbane, Australia