



This is a repository copy of *Tactical interventions in online hate speech : the case of #stopIslam*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/176538/>

Version: Published Version

Article:

Poole, E., Giraud, E.H. and de Quincey, E. (2021) Tactical interventions in online hate speech : the case of #stopIslam. *New Media & Society*, 23 (6). pp. 1415-1442. ISSN 1461-4448

<https://doi.org/10.1177/1461444820903319>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Article

Tactical interventions in online hate speech: The case of #stopIslam

new media & society
2021, Vol. 23(6) 1415–1442
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1461444820903319
journals.sagepub.com/home/nms



Elizabeth Poole, Eva Haifa Giraud 
and Ed de Quincey

Keele University, UK

Abstract

This article sets out findings from a project focused on #stopIslam, a hashtag that gained prominence following the Brussels terror attack of 2016. We initially outline a big data analysis which shows how counter-narratives – criticizing #stopIslam – momentarily subverted negative news reporting of Muslims. The rest of the article details qualitative findings that complicate this initial positive picture. We set out key tactics engaged in by right-wing actors, self-identified Muslim users, would-be allies and celebrities and elucidate how these tactics were instrumental in the direction, dynamics and legacies of the hashtag. We argue that the tactical interventions of tightly bound networks of right-wing actors, as well as the structural constraints of the platform, not only undermined the longevity and coherence of the counter-narratives but subtly modulated the affordances of Twitter in ways that enabled these users to extend their voice outwards, reinforcing long-standing representational inequalities in the process.

Keywords

Counter-narratives, hate speech, Islamophobia, right-wing populism, social media, Twitter

Introduction

In the wake of the well-documented rise of populist right-wing politics in Europe and North America (Emcke, 2019; Kellner, 2016), there has been widespread concern about social media being used in ways that normalizes xenophobia (Evolvi, 2018; Feshami, 2018; Siapera, 2019) and propagates disinformation about minority groups (Farkas et al.,

Corresponding author:

Elizabeth Poole, Keele University, Keele, Staffordshire ST5 5BG, UK.
Email: e.a.poole@keele.ac.uk

2017; Horsti, 2017). Glimmers of hope, however, have also been argued to exist, with the same media that have enabled the rapid spread of hate speech also appearing to offer opportunities to contest it. Platforms such as Twitter, for instance, have been seen to enable the formation of anti-racist and anti-Islamophobic counter-narratives that gain circulation in the wider public sphere (Dawes, 2017; Jackson and Foucault Welles, 2015). To draw on Sarah Jackson and Brooke Foucault Welles's (2015) definition, these online counter-narratives consist of 'outward-looking attempt to challenge mainstream narratives' (p. 4), which 'reframe and retell' (p. 5) existing depictions of an issue. In the process, counter-narratives can enable perspectives that are often excluded from the mainstream media (i.e. television and national newspapers) to be 'articulated en mass' (Jackson and Foucault Welles, 2015: 4).

In this article, we conceptualize the significance of tactics that were engaged in by different actors involved in circulating #stopIslam: a hashtag that initially trended on Twitter following the Brussels terror attacks of 2016. The hashtag seemed to reflect both long-standing anti-Islamic discourses in mainstream Western European media (Baker et al., 2013; Poole, 2002) and the contemporary intensification and internationalization of these discourses, typified by campaigning surrounding Brexit and the US presidential election, which saw an 'insiders'/'outsiders' discourse become a central component of mainstream political campaigning across Europe and North America (Martin, 2016; Modood, 2017). However, what was notable about #stopIslam is that the reason it trended on Twitter was not due to people using it to spread hate speech. Instead #stopIslam grew to prominence because those critical of Islamophobia were engaging with it in order to condemn its original sentiment. As such, the dynamics of the hashtag offer insight into the processes through which particular social media platforms can serve both as vectors for the spread of hate speech and create scope to contest it.

The article evaluates the efficacy of tactical attempts to contest #stopIslam, with a particular focus on the limitations of these tactics in light of well-organized opposition on the part of actors seeking to perpetuate its original anti-Islamic sentiment. These findings are primarily based on qualitative data about who was engaging with #stopIslam, how the hashtag was being deployed and the dialogue that surrounded its deployment. Although our focus here is on qualitative research, the data we are drawing on is derived from a larger mixed-methods project about key potentials and challenges facing online counter-narratives against racialized Islamophobic hate speech (see Poole et al., 2019). This big data study gathered and analysed all tweets using #stopIslam for 40 days following the Brussels terrorist attack (22 March 2016), using computational methods. The findings were further interrogated through a manual quantitative analysis of the 5000 most shared tweets and finally a qualitative analysis of the top 150 retweets and their associated comments. It is these qualitative materials that we focus on here.

After setting the scene by providing a brief summary of our quantitative data and the methodology used to analyse the qualitative materials that are the focus of this article, the main body of the article delineates dominant tactics that were used, first, by self-identified right-wing Twitter users (to propagate the hashtag), before turning to would-be allies and self-defined Muslim users (who attempted to contest it). In addition to evaluating the significance of these tactics in shaping the overarching discourse associated with #stopIslam, we also elaborate on their theoretical significance and elucidate how particular tactics

productively complicate the distinction between tactics and strategies (as made by De Certeau, 1984) that is often used to understand public engagement with social media (e.g. De Ridder, 2013; Liao and Humphreys, 2014; Manovich, 2009). We suggest the approaches used by key actors instead serve as ‘tactical interventions’ (see, Giraud, 2018, 2019), wherein users attempt ‘to interfere in complex communication ecologies by modulating the affordances of particular media, a sort of digital weapon of the weak intended to counteract the growing power differentials in this realm’ (Lezaun, 2018: 224). What events surrounding #stopIslam elucidate, however, is that – in the context of commercial social media – though particular tactics might subtly modulate the affordances of social media, these shifts often work to intensify rather than unsettle existing representational inequalities.

Literature review

Recent research exploring how social media lends itself to reactionary politics (e.g. Schradie, 2019) resonates with a longer tradition of theoretical critique, wherein commercial media platforms have been conceptualized as undermining liberal–democratic politics by fostering a: ‘political-medialogical setting [. . .] of dissensus, incredulity and competing conceptions of reality’ (Dean, 2009: 147). At the same time, it is striking that – as recently as 2016 – researchers were pointing to a lack of *empirical* research about the role of digital media in right-wing movements, in comparison with the large body of work about progressivist groups (Mercea et al., 2016: 285). Although this gap in knowledge is being rapidly redressed (e.g. Ouellette and Banet-Weiser, 2018; Schradie, 2019), this is an ongoing project and further work needs to be undertaken to develop theoretically informed empirical work, which conceptualizes the dynamics of communication ecologies that allow hate speech to circulate and become normalized.

Digital media technologies, for instance, have often been connected with right-wing populism in generalized ways. Popular commentaries on the spread of mis- and disinformation have argued that the Internet has ‘been the all-important, primary, indispensable engine of Post-Truth’ (D’Ancona, 2017: 49). The affordances of platforms such as Twitter have been argued to create ‘ideological silos’ and ‘powerful echo chambers of misinformation’ that have displaced mainstream news media and lent support to narratives propagated by the populist right (Ott, 2017: 65). Although often containing valuable insight, such commentaries nonetheless tend to construct overly deterministic narratives about the role of particular technologies, enacting what Emiliano Treré (2019) describes as the ‘one-medium fallacy’ (p. 9) that neglects the way the affordances of any given platform are shaped by their position within a broader communications ecology (composed of interactions between different media platforms and communicative practices).

Cognizant of the danger of privileging a particular platform in isolation, without situating it in its constitutive communications ecology, it is *because* of Twitter’s relationship with a range of other media platforms and communicative practices that the dynamics of hashtags such as #stopIslam are significant. Due to broader declines in resources for investigative journalism within the mainstream media (Fenton, 2010), Twitter has been increasingly treated as a news source in its own right (Broersma and Graham, 2013). The relationship between the platform and the mainstream media has given Twitter particular significance in relation to anti-racist politics.

Existing research into the relationship between Twitter and the media has pointed to the way hashtag campaigns have created visibility for anti-racist narratives (Rambukanna, 2015). Although the temporary, 'ad hoc' (Bruns and Burgess, 2011; Dawes, 2017) publics who mobilize around particular hashtags are often transitory in themselves, they can have a lasting impact on mainstream media discourse. As Jackson and Foucault Welles (2015, 2016) elucidate: in the context of campaigns such as #BlackLivesMatter and #Ferguson, hashtags were not only used in ways that gave online visibility to perspectives that were formerly marginalized within the media, but gave platforms to community spokespeople and offered frames that went on to inform mainstream news.

Yet, although Twitter has been used to anti-racist ends, it has equally been accused of propagating everyday racism that ranges from the circulation of gifs and memes that perpetuate stereotypes (Sharma, 2013) to more strategic uses of humour to disguise racism and misogyny that has been a hallmark of the right-wing populism (Ringrose, 2018). These latter developments are entangled with more organized forms of racism, associated with what Wahl-Jorgensen (2019) describes as 'angry populism', an 'emotional regime' that creates 'exclusionary solidarities' by portraying minority groups as shared enemies (p. 110). While Twitter itself is argued to provide a space for 'collective affect' (Abdel-Fadil, 2019) where exclusionary solidarities are reinscribed, this consolidation of identity often occurs through the relationship *between* Twitter and other media. A growing body of research indicates the platform's role in disseminating discriminatory rhetoric that originated in the echo chambers of right-wing web fora (Siapera, 2019), with groups tactically using 'trigger events' (Awan, 2014; Copsey et al., 2013) such as the Brussels attacks – and the hashtags associated with these events – to gain wider visibility for their views. These dynamics mean that despite features such as hashtags bringing those with different ideological commitments together, interactions on Twitter have not been found to serve a Habermasian public sphere ideal of debate but result in antagonistic encounters (Evolvi, 2018) that serve to entrench reactionary views (Karlsen et al., 2017).

Methods and sample

Building on these debates, in order to analyse the framing and dynamics of #stopIslam, we adopted a multi-method approach incorporating computational analysis with manual quantitative and qualitative content analysis (Cresswell and Clark, 2007). The first approach, often known as the study of 'big data', has been adopted for many critical data studies that examine the representation and use by marginal groups to highlight issues around identity politics (including Black identity politics in the United States, Jackson and Foucault Welles, 2015, 2016; work on refugee-related hashtags, Siapera et al., 2018; and the #metoo movement, Clark-Parsons, 2019; Mendes et al., 2018). These methods are extremely beneficial for analysing big data sets; they can reveal the longitudinal patterns in the development of networks and the framing of a particular issue, allowing for the categorization of data and thus reduce time manually coding. However, they have their limitations, with the main criticisms relating to the assumed accuracy, transparency and objectivity in the way the results are gathered and presented (boyd and Crawford, 2012). Similar criticisms have been

levelled at quantitative content analysis, so triangulation is the best approach to address the limitations in any methods used, with a critical eye on research design and reflexivity when analysing and presenting the results.

We used Twitter's enterprise API (Application Programme Interface) platform Gnip¹ to ensure the full data set was collected, including all tweets using the hashtag #stopIslam from just before the attacks (on 22 March 2016) and for 40 days following it (20 March–29 April 2016).² After removing the spam from the 551,400 tweets we received from Twitter, we were left with 66,764 unique tweets and 235,578 retweets (shared original tweets), 302,342 in total. As well as applying content and descriptive analytics, we also undertook a network analysis of those users who had retweeted others and been retweeted. This was followed by a manual quantitative analysis of the 5000 most shared tweets. After sorting for deleted accounts, this left us with 4263 tweets. We used a coding schedule to measure the date and time of tweets, location, gender, religion and political and/or institutional affiliation of the tweeter, topic of the tweet and to establish whether they were part of the dominant narrative (against Muslims) or counter-narrative (that contested the original negative narrative). This enabled us to verify the results of the computational analysis; we were particularly careful about not making assumptions about identity (religion, for example) and location and coded these as unknown, unless they were specifically identified by the account users (using the biographies of account holders to verify these details). For a longer discussion of methods, particularly relating to the quantitative stages, please refer to Poole et al. (2019).

Qualitative content analysis allows for a richer analysis of the structure, language, imagery and interactions of participants, an approach that (in line with Deacon et al., 2007) can be valuable in establishing the construction of activist narratives and identity. We analysed the 150 most shared tweets with three coders analysing 20 each (two of these being the authors who had a strong insight into the data from previous stages of the project). To ensure inter-coder reliability, there were three meetings at key moments throughout the analysis, before, during and after the process. At each meeting, a sample of tweets (10) was co-analysed to resolve any discrepancies and ensure a consistency of interpretation. Written guidelines were also issued prior to these meetings and refined if required in subsequent discussions. Two of the authors then analysed the findings across the data set, rather than focusing on their own 50 tweets, which provided another level of verification.

The comments we examined were highly polarized; posts tended to either directly challenge rhetoric in the tweets they were responding to or overtly praise it, which made them relatively straight forward to categorize consistently between coders. A popular counter-narrative tweet in our sample, for instance, was a meme depicting the Ku Klux Klan (KKK), which pointed to the hypocrisy of people focusing on Islamic terrorism and neglecting White supremacism. We labelled messages that disputed the factual content of the tweet as 'disagreeing' (e.g. tweets that argue the KKK 'hasn't done anything recently', 'don't have a state', 'are less of a global threat', or posted links to articles about Islamic extremism as a counter-point). In contrast, we categorized comments as 'agreeing' if they praised the content of the tweet (e.g. by stating it showed the hashtag was trending for the 'right reasons' or linked to articles about right-wing extremism that backed up its sentiment). More nuanced messages, which did not adopt an overt,

polarized, stance, were sparse. A small number of exceptions to this rule came from Muslims themselves, who made critical comments about terrorism (using terminology such as ‘extremism’ and ‘radicalization’) but situated these comments as part of a broader defence of ‘ordinary Muslims’ and condemned the overall sentiment of #stopIslam, which led to us categorizing such comments as ‘disagreeing’ with tweets propagating the original narrative and ‘agreeing’ with counter-narrative sentiment (see section ‘tactics engaged in by Muslim users’, p. 18).

Tweets that were more difficult to categorize, such as potentially sarcastic messages, were also discussed in coding meetings. Sarcasm can be difficult to detect online – indeed it is often exploited to deny that particular statements are hate speech (Frenda, 2018) – so, for the purpose of our sample, we defined sarcasm as tweets that seemed to infer one thing but used explicit signifiers to show they meant the opposite. For instance, one of the most widely circulated tweets in our sample appeared to be a positive statement about Islam being a peaceful religion, but this was directly undermined by a meme underneath the tweet that associated Islam with violence. Due to tweets like this having clear signifiers of their political stance, we found that they had been categorized consistently between coders. The prominence of such tweets, moreover, elucidates the value of the qualitative approach used for this study, as it provided the necessary contextual detail to interpret the meaning of tweets in ways that could not be captured by quantitative data.

In order to adhere with recent ethical guidance on the handling of social media data (e.g. Townsend and Wallace, 2016), we only include data that have been processed in ways that do not identify individual users, mostly by describing the content of tweets or citing fragments that have generic wording (instead of quoting verbatim) and reproducing memes that were shared by multiple users.

Overall, using this triangulated approach ensured that we could capture a comprehensive and robust picture of the dynamics within this hashtag; in this article, focusing on the tactics of participants in their interactions with each other, which resulted in the framing of the hashtag in a particular way.

The dynamics of a racist hashtag: quantitative data

The big data analysis suggested the negativity of the hashtag with a higher proportion of keywords having a negative slant such as ‘stop’, ‘ignorant’, ‘terrorism’, ‘hate’ or ‘stupid’ (see Poole et al., 2019 for a more extended discussion of the quantitative findings). A search of top related hashtags consolidated this finding and revealed the politics of those circulating the hashtag, linking its ongoing circulation to US conservative groups on Twitter (#tcot) and right-wing political discourse (#wakeupamerica) in the build up to the election of Donald Trump (#Trump2016). User bios and location data confirmed that many of those circulating the hashtag were US based. The dominance of users from the United States reflects a media landscape where political events on social media are increasingly being used opportunistically by political groups. In this instance, though the event was in Europe, it was being utilized by right-wing individuals and communities in the United States to leverage support for Donald Trump’s 2016 electoral campaign. According to Modood (2017), Islamophobia has been central to the rhetoric of Trumpism

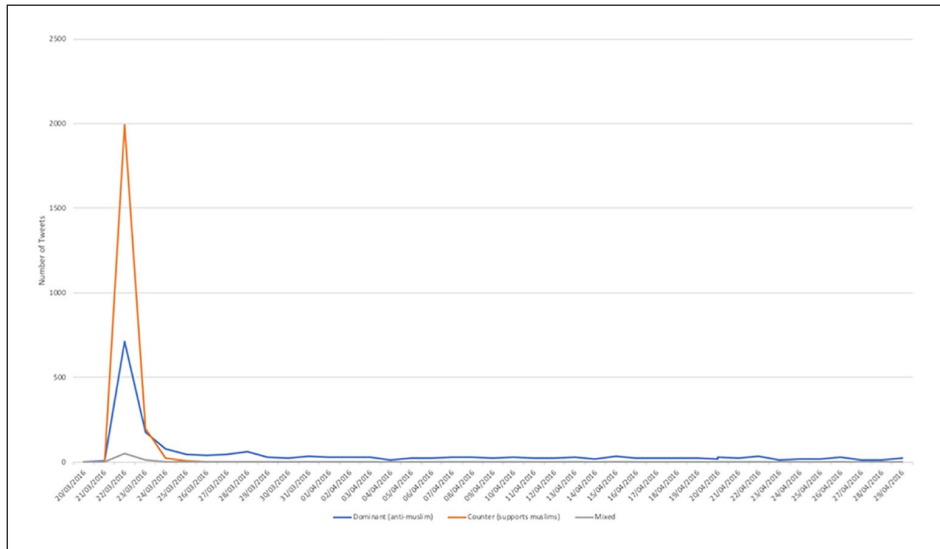


Figure 1. Timeline of Tweets.

as a ‘nationalist-populist’ movement which requires the identification of ‘insiders and outsiders’ to garner support (for more discussion of the connection between this hashtag, post-truth, and right-wing populism and for further elucidation of these dynamics, see Poole et al., 2019). Furthermore, a network analysis of users retweeting each other revealed the close ties of the right with particular accounts serving as nodes who anchored these networks.

When we examined the most *shared* tweets, however, these were predominantly counter-narratives supporting Muslims, including nine out of the top 10 retweets. It was then evident that many of the negative words being used in the tweets were attacking the hashtag rather than Muslims as in ‘why is this hateful hashtag trending?’ However, despite the largest proportion of tweets in our sample supporting the counter-narrative, this was relatively short-lived. Examining the timeline of the hashtag showed that counter-narratives were more likely to be posted within 24 hours of the event, while those attacking Muslims were more persistent over time (Figure 1). The longevity of the anti-Muslim discourse was also evident in subsequent occurrences of the hashtag. After the Manchester and London terrorist attacks, 2017, it had reverted back to being wholly negative and anti-Muslim. Also of significance was the number of media outlets reporting on the hashtag trending that overwhelmingly focused on the counter-narrative (discussed in more depth below). In this way, counter-narratives about Muslims were momentarily able to subvert dominant (negative) news about Muslims (a framing that is becoming more common in light of the growing prominence of right-wing extremism). These patterns will now be further explored through a discussion of the findings of the qualitative analysis of the most shared tweets.

Dominant	Counter-narrative	Mixed	Missing	Total
41 (27%)	91 (61%)	3 (2%)	15 (10%)	150 (100%)

Figure 2. Position taken in the 150 most shared tweets.

Dominant	Counter-narrative	Mixed	Missing/Unknown	Total
97 (65%)	18 (12%)	13 (9%)	22 (15%)	150 (100%)

Figure 3. Position taken in the comments to the most shared tweets.

Overview of tactical engagements with #stopIslam

Of the 150 most shared tweets, most constructed a counter-narrative against #stopIslam (Figure 2), 29 of these 91 tweets were posted by accounts identifying as Muslim (19%). Yet, when we examine the comments, it is clear that *responses* to these tweets were predominantly in line with the sentiment of the original hashtag (Figure 3). This dynamic – of the majority of tweets contesting anti-Islamic sentiment, while the majority of comments propagated it – was the most common pattern of how the hashtag initially played out. The ‘trench warfare’ dynamic described by Karlsen et al. (2017) was reflected in the minimal interactions that arose between participants in the comments, with almost no interaction between the original tweeters and the respondents to their tweets.

The most shared tweets (perhaps inevitably) tended to be those that had the most followers: opinion leaders, activists, celebrities (who we defined as those whose fame derived from a context other than their social media use) and influencers (whose fame originated due to their social media prominence). Interestingly, in the time from when we initially collected the data to writing up our analysis, Twitter had removed a large proportion of ‘offensive’ content, so, for example, where initially the most shared celebrity tweet was retweeted 3904 times, had 6254 likes and 362 comments, at the time of writing, it had 3676 retweets, 5883 likes and 337 comments. This was common across the tweets and demonstrates Twitter’s role as an ‘active agent’ (Kriess and McGregor, 2017) in the direction and legacy of the narrative. Not surprisingly, the number of comments correlated with the number of times tweets were shared (Figure 4).

Also of note was that the 150 most shared tweets were generated by only 110 users; this meant that although the majority of tweets derived from unique accounts, a small number were from individuals posting multiple times, with 14 users appearing more than twice. One user’s tweets, however, appeared 21 times in the sample, suggesting that even though the overall *proportion* of users whose tweets appeared multiple times was small, these users could play a significant role in propagating particular narratives (indeed, as discussed in more depth below, this particular user was significant in anchoring key right-wing user networks that surrounded the hashtag).

Tweets	No of Retweets	No of comments
1-50	1000-2000	Up to 400
51-100	400-600	40-100
101-150	200-300	10-40

Figure 4. Average number of retweets and comments: most shared tweets.

Topic	Number	%
Negates the relationship between Islam and Terrorism	1,419	33.3
Islamification/spread of Islam	967	22.7
Islam as a negative force	704	16.5
Islam as a positive force	397	9.3
Anti-left agendas	209	4.9
Muslims as victims/discrimination/Islamophobia	141	3.3
Reflecting on the # trending	127	3.0
Anti-far right discourse	22	0.5
Other	55	1.3
Total	4,041	94.8
Other identified themes	222	5.2
Total Sample	4,263	100

Figure 5. Themes of tweets and retweets.

Discursive frames

Figure 5 shows the dominant themes of coverage we identified through our quantitative content analysis of 4263 tweets. While the most prominent single topic of tweets appears to be those that negate the relationship between Islam and terrorism, when topics are combined, there is a fairly even split between positive (46.4%) and negative (44.1%) discourses about Islam. At the time, we speculated that this could be because there was less diversity in the counter-narratives being shared. However, the qualitative data demonstrated that there was a greater range in the content of counter-narrative tweets and comments than among the dominant narratives.

Dominant narratives tended to focus quite singularly on equating Islam with terrorism and ‘evil’, using decontextualized extracts from the Quran to support these arguments, and linking the hashtag with anti-Democrat agendas. Other topics paralleled that of mainstream media discourse (in an exaggerated form) relating refugees and immigrants to radicalization, and constructing narratives of Islamification and thus providing a coherent and predominantly negative story about Islam (cf. Siapera et al., 2015).

Echoing other work that has foregrounded how frames established by negative narratives constrain attempts at contestation (Siapera et al., 2018), sentiments contained in the original hashtag also structured the counter-narrative content. For example, as depicted in Figure 5, many of the tweets reacted defensively towards the original narrative's attack on Islam and Muslims, and were critical of the hashtag as ignorant or Islamophobic, an approach that was easily undermined by apparently 'factual' information offered by right-wing users (such as statistics, links to news stories and memes mocking positive sentiment). In the following sections, we discuss the tactics demonstrated within the most common categories of tweet including anti-Islam tweets, counter-narrative tweets, tweets by Muslim participants and tweets shared by celebrities and media organizations.

Tactics of the right: appropriation, affirmation and humour

The populist political right, as represented in our sample, are a diversity of participants ranging from self-identifying conservatives (many also as Christians) to nationalists and White supremacists, some identifying as the 'new' or 'alt-right'. Combinations of patriotic and Christian symbols were used to signify these intersecting identities, suggesting that religious iconography was being used less to signify spiritual identity and instead to denote *political* affiliation (Modood, 2017). While the participants did not tend to belong to organized groups, and predominantly engaged with the hashtag as individuals, their activities on Twitter – such as liking and retweeting – created a tightly bound network of like-minded individuals. These connections were made especially visible on conducting a network analysis that depicted which users retweeted one another most frequently. As illustrated by Figure 6, those circulating tweets that conformed to the narrative's original, anti-Islamic sentiment (clustered here on the left-hand side of the image, oriented around some particularly prominent accounts signalled by the large nodes) consisted of users who retweeted one another multiple times. In contrast, users engaged with the counter-narrative formed a much more dispersed network.

These close-knit connections between right-wing users enabled the rapid spread of memes that parodied the counter-narrative and enabled these actors to work together to close down counter-narrative intervention, by tagging more high profile users, for example. It was these 'opinion leaders' or 'influencers', journalists or vloggers of right-wing content that had the most shared tweets. The most prolific of these was the account of a prominent right-wing blogger, who described herself as 'anti-Islam'. She tweeted #stopIslam 93 times in total, was responsible for 21 of the most shared tweets (as described above), and was often strategically tagged into tweets by other users, a tactic that was used repeatedly as means of introducing the hashtag to a global audience. Many of the high-profile participants (such as this account), for instance, only began tweeting later in the day or day after the attacks, after being tagged by other users, yet their posts still had some of the highest numbers of shares. These dynamics reflect how the affordances of platforms such as Twitter reinforce the dominance of 'elites' and hence the power structures present in traditional media. The influence of these highly visible users is illustrated by the only negative tweet appearing in the top 10 most shared tweets (ninth position) being circulated by a 'new right' vlogger with 974,000 followers. The tweet initially

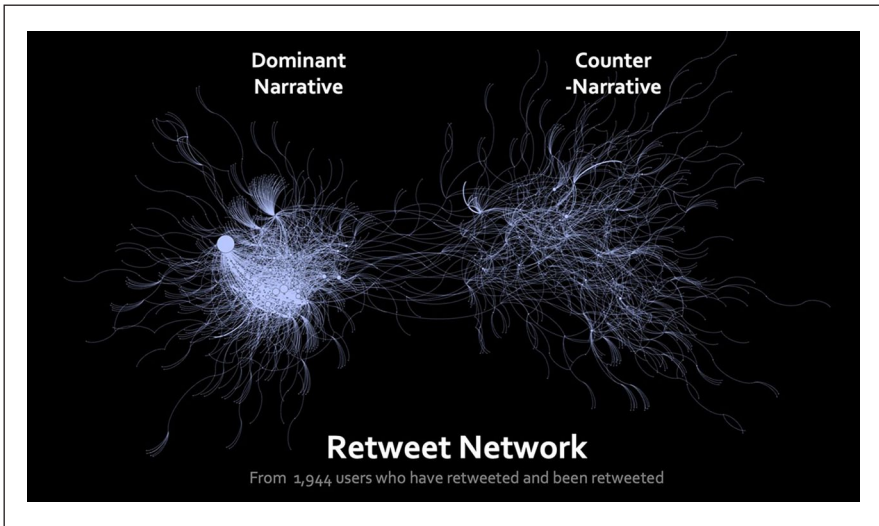


Figure 6. Retweet network.

suggests that #stopIslam is Islamophobic and Islam is a religion of peace but the meme demonstrates its sarcasm, providing a graph claiming to be from Wikipedia (though the image does not currently appear on the site) that positions Islamism as the most prolific architect of terrorism (Figure 7).

The simple tactic of affirmation through volume was a characteristic response to tweets such as this; of the 116 comments, for instance, 63 agreed and 19 disagreed with the content of the tweet. Those agreeing also used sarcasm and humour to make their point ('that religion of peace just killed . . .') and shared memes including multiple decontextualized quotes from the Quran to illustrate their point. While the counter-narratives tried to offer corrections to the information presented in the chart, disputing the categorizations, definitions and providing opposing quotes from the Quran, as discussed in more depth below, the volume and persistence of the right-wing tweeters proved hard to contest.

This emotional investment by anti-Islam participants is further demonstrated by the response to counter-narrative tweets or comments by those defending Islam, which we have termed 'flak' (following Herman and Chomsky, 1988). As well as the high volume of comments agreeing with Islamophobic tweets, and deployment of memes and statistics, participants used the tactic of repeating statements and engaging in trolling across different users' accounts. For example, there were frequent instances of the same people commenting on multiple tweets, with statements such as 'Muslims are terrorists: fact'. In one example, a self-identified Muslim posted a quote from the Quran which suggested that those circulating the hashtag should educate themselves (Figure 8), which they followed, in the comments, by, and unusually, a more lengthy engagement to defend Islam. In a prolonged discussion, where the Muslim user attempted to challenge the source of the anti-Islam actor's knowledge, the latter was rescued by allies posting memes of page

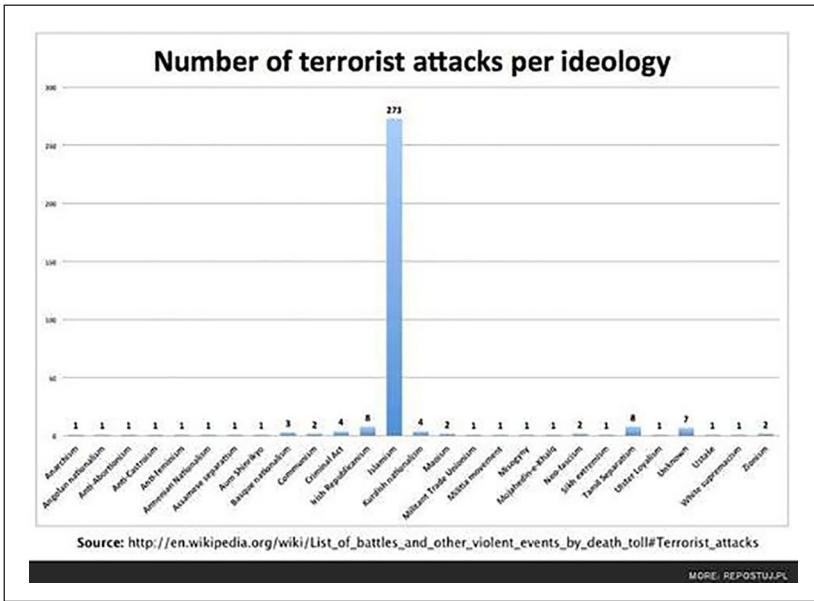


Figure 7. Meme shared in the most retweeted ‘dominant narrative’ tweet.

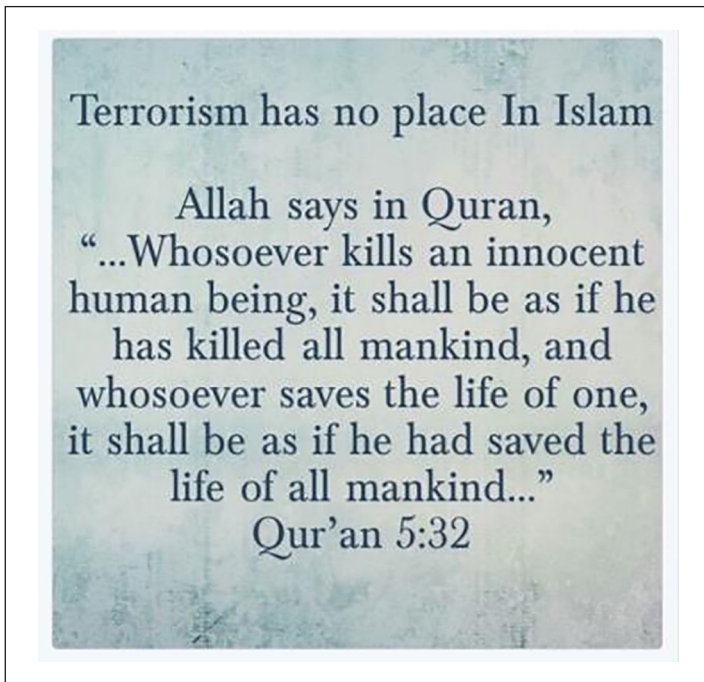


Figure 8. Meme shared by Muslim user: Counter-narrative.

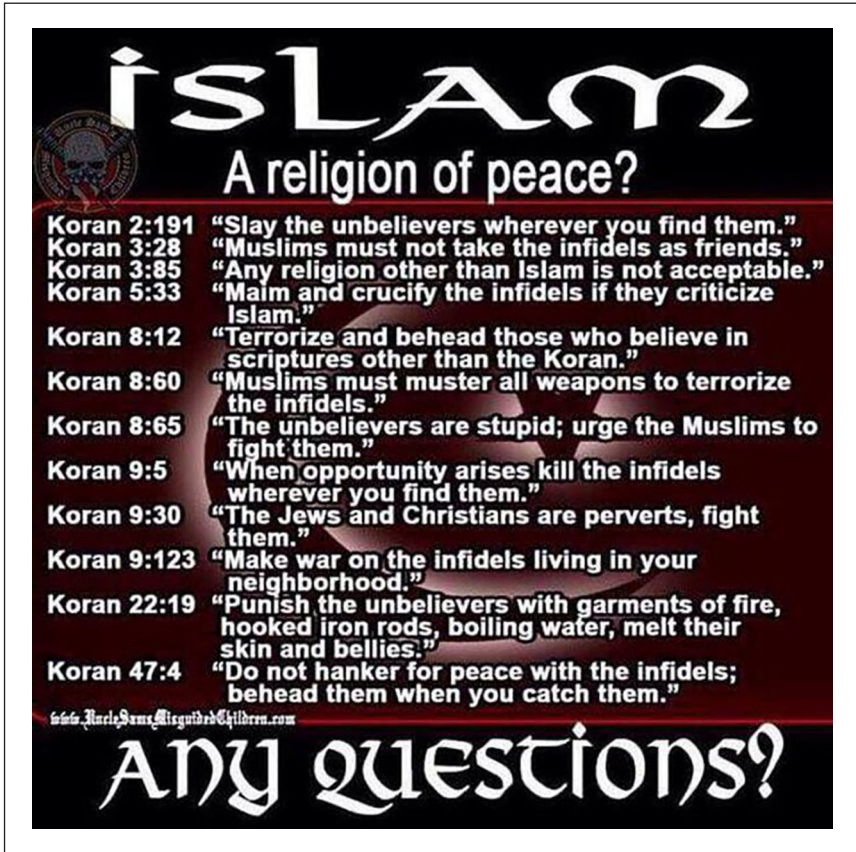


Figure 9. Meme used to challenge the counter-narrative in Figure 8.

references to the Quran that appeared to corroborate their claims (Figure 9). Here the volume of responses worked to make any further counter-narrative intervention appear futile, silencing the critic and increasing the legitimacy of these views.

Additional tactics of the right, in response to counter-narrative tweets, include disputing, dismissing and refuting the claims, and defending their own actions. For example, a counter-narrative tweet, accompanied by the meme shown in Figure 10, refutes any link between Islam and terrorism and questions the relative value of White and ‘Muslim/ Arab’ people. Responses to this counter-narrative included *disputing* using ‘alternative facts’ (Quranic quotes, statistics, memes and links to articles); *dismissing* the importance of Muslim lives (‘I don’t care if you kill each other’); *refuting* the claim (e.g. ‘we did pray for them’); and *defending* their actions (‘It’s not irrational to care more about attacks on home than other attacks’). Further examples of attempts to dispute the claim in Figure 10 included arguing Belgium is multicultural and no longer identifies as a ‘White country’, offering ‘facts’ about European aid provided to Syria, and circulating memes such as the one shown in Figure 7. Also evident were White supremacist conspiracy theories,



Figure 10. Counter-narrative meme.

which accused ‘liberals’, Muslims and the mainstream media of themselves propagating fake news and deleting right-wing content that revealed the ‘truth’. Some of these tactics were evident in comments beneath counter-narrative tweets that accuse ‘the West’ of hypocrisy in their response to terrorism.

The use of appropriation whereby a separate event, reported elsewhere, is used to provide additional evidence to support the hegemonic discourse was also evident. In this case (Figure 11), the source (*The Daily Mail*) provided additional authority, an authority further legitimized by the fact it has been left in place by Twitter even though a significant proportion of other openly anti-Islamic content has now been deleted. In this tweet, the author selects and highlights key factual evidence from the *Daily Mail* article, how many times the victim was stabbed, which is accompanied by the tagline ‘Murdered by the cult of peace’ and linked to several hashtags including #Trump and #NoRapefugees incorporating several right-wing themes (Islamophobia, anti-immigration).

This article was posted by a user self-identifying as a ‘pro-life Catholic’ and resulted in total agreement in the comments. Siopera (2019) has previously noted how semi-organized right-wing groups link new events (often circulated by mainstream media) to pre-existing hashtags which are circulated through influential accounts, an example of ‘transnational contagion’, and demonstrates the ‘instrumentalized’ use of Twitter for political purposes. Connecting these hashtags to particularly controversial events like terror attacks triggers the kind of affective response that is needed to give populist narratives traction. It is clear how anger or outrage is ideologically and discursively constructed here to maintain hegemony.



Figure 11. Meme derived from a mainstream news source.

This leads us to the final tactic of note which is the instrumental use of anti-Islam content as political propaganda to bolster Trump’s campaign for the White House. Quotes from Hillary Clinton supporting Muslims and conspiracy around Obama’s heritage are used to question their judgement – ‘they will bring in terrorists’; the fact that this support is espoused by the political enemy in turn further demonizes Muslims to right-wing audiences.

Tactics from would-be allies: generalized criticism of hate speech

While uses of #stopIslam that received the highest number of retweets countered the hashtag’s original meaning, the way these messages attempted to combat Islamophobia generated tensions that undercut their aims. In this section, we detail a series of tensions that emerged and undermined tweets that tried to criticize the hashtag, beginning with a discussion of would-be allies who sought to defend Muslims.

One of the most common tactics used by those who sought to contest #stopIslam was making generalized criticisms of hate speech. Terrorism was repeatedly asserted to have ‘no religion’, while Islam was defended on the basis of being a ‘religion of peace’, and people’s positive relationships with Muslim friends and colleagues were emphasized in order to debunk stereotypes (one of the most shared tweets described Muslim neighbours regularly bringing food to a particular user when they were ill). Memes were frequently shared to underline these sentiments, ranging from aforementioned images of the KKK (which were used to frame Islamophobic rhetoric as hypocritical) to a clip

of Ben Affleck defending Islam on HBO's *Real Time* (often circulated with reference to Batman's heroism).

The hashtag itself was also criticized for being Islamophobic, or propagating hate through 'ignorance' and 'fearmongering', with some users directly connecting the spread of #stopIslam with White supremacy. Others attempted to create a more nuanced picture of the reasons behind terror attacks seeking to reframe the problem using alternative hashtags (such as #stopISIS), and situating the events in a geopolitical context where it is not religion but the legacies of colonialism and ongoing racism that are framed as fostering alienation and violence.

As we detail in more depth below, due to the widespread circulation of these tweets (the most widely shared tweet – which questioned why people were focusing on Islam in light of rising violence associated with White supremacy – for instance was shared 6643 times), they were valuable in affording the counter-narrative visibility within the mainstream media. At the same time, these tweets tended to rely on making generalized assertions, which meant they left openings for self-identified, right-wing Twitter users to undermine them with more specific counter-evidence. As described above, due to the 'evidence-based' appearance of much of the commentary propagated by the right and the frequency with which it was posted (with some tweets receiving over 300 negative comments), the counter-narrative sometimes, paradoxically, contributed to the circulation of hate speech rather than contesting it.

These dynamics were evident in the comments to a tweet that questioned why so much attention was focused on Islamic terrorism, in light of recent mass shootings in the United States by those affiliated with the Christian right. Of 37 comments, only two were supportive of this tweet, while the rest criticized the link between Christianity and violence; numerous comments offered statistics to debunk this connection and several inferred the poster themselves was racist in associating Christianity with Whiteness. One comment was especially notable, taking the form of a question that asked whether all of those trying to challenge the hashtag were working from 'the same script repeated over and over again'.

Ironically, it was those propagating negative comments whose activities were more akin to a 'script'. As illustrated by Siapera (2019), extreme right online communities routinely offer (highly prescriptive) guidance for propagating hate speech effectively on social media. As detailed within an infamous *New Yorker* piece on the style guide of neo-Nazi site the *Daily Stormer*, for those posting on social media, users are instructed to include: 'as much visual stimulation as possible', to 'appeal to the ADHD culture', while passages from mainstream sources must be unaltered, so that 'we can never be accused of "fake news" – or delisted by Facebook as such' (in Marantz, 2018). Echoing academic research that has foregrounded the tactical role of humour and sarcasm (Frenda, 2018), such approaches are encouraged as the 'unindoctrinated should not be able to tell if we are joking or not' (in Marantz, 2018) in order to avoid censure while contributing to 'dehumaniz[ing]' rhetoric. These guidelines resonate with our analysis of negative comments that attacked the tweets of would-be allies, in which users drew on a fairly narrow and prescribed repertoire of action (oriented around the circulation of 'humorous' memes, statistics and hyperlinks).

While it was not necessarily the case that all of the tweets we analysed followed a script in the purposive or prescriptive sense described above, the tightly bound nature of right-wing networks enabled the circulation of particular memes and phrasing nonetheless afforded these comments a sense of uniformity. A commonplace meme, for instance, took the form of a bingo card and displayed a series of boxes that detailed commonplace attempts to contest associations between Islam and terrorism – including ‘terrorists are not true Muslims’, ‘you took that verse out of context’ and ‘you are Islamophobic’ – that were juxtaposed with racialized stereotypes, in an attempt to mock common criticisms of Islamophobia while simultaneously propagating hate speech. The particular discourse employed in negative comments also contained recurring motifs that have come to be associated with the right (including ‘social justice warriors’ [SJWs], ‘snowflakes’ and references to ‘virtue signalling’), to dismiss critics.

Again the *frequency* with which users defending hate speech were attacked in comments underneath their tweets also worked to shut down any sense of a counter-narrative. As with the above examples, of all of the tweets we examined, negative comments dominated in every instance (with the exception of celebrity accounts, see below). Even those whose tweets had slightly lower levels of engagement expressed surprise and anxiety over criticism they received. A user who ordinarily posted about fashion and music for instance, repeated the common refrain that terrorism, rather than Islam, was the problem and received 328 retweets and 30 comments. After initially trying to refute an opening wave of comments containing Islamophobic rhetoric, this user stated they were no longer going to participate in discussion due to emotional exhaustion.

What is hinted at in this instance is the level of emotional labour involved in participating in narratives against hate, in a media environment that has enabled tightly bound groups of individuals both to rapidly respond to tweets and who have a ready repertoire of tactics and resources to draw on that require active work to contest. Questions need to be asked, however, about *who* is able to disengage in this way. Our findings point to the relative ease by which would-be allies can shield themselves, that is, by deleting tweets, dropping out or disengaging. In contrast, those identifying as Muslim or who had obvious racial identifiers on their profile would have to actively hide, obscure or deny an aspect of their identity in order to shield themselves from emotional labour in this way. Of course, our methodology could result in Muslims falling into the ‘allies’ category if they have not identified as such explicitly. Indeed, the act of not calling attention to this aspect of identity may be significant and, while the reasons for this cannot be assumed here, could be an important area of future research.

Social media advocacy campaigns have long been accused of inadvertently perpetuating neo-colonial narratives and ethnocentric stereotypes in their attempts to speak for others (Holohan, 2019; Maxfield, 2015; Poole et al., 2019; Torchin, 2016). While the counter-narrative against #stopIslam might not have overtly perpetuated stereotypes, it was nonetheless in danger of reinscribing other inequalities. Tweets of would-be allies illustrate how counter-narratives themselves can be ‘hijacked’ in order to propagate the very hate speech these narratives were contesting. If these users then ‘drop out’ in ways that are not available to those most affected by Islamophobia, then their tweets could inadvertently contribute to the circulation of hate speech while being able to avoid its consequences.

Tactics employed by Muslim Twitter users: contesting stereotypes

Although there are similarities in the tweets and tactics of Muslims and would-be allies, it is important to analyse these separately due to the specific attack on this aspect of identity. In this section, we discuss findings solely related to tweets where the user clearly identified as Muslim (in their bio) to avoid making assumptions based on other characteristics. In general terms, (visible) Muslim voices were evident in this data set but not to the extent we might expect (15.8% of 4263 retweets). Only 11 of the 50 most shared tweets were posted by Muslims, just two in the top 10. Both of these were social activists from Pakistan. Most of these tweets received more flak than support and hence there was little engagement from Muslims in the comments.

The most prevalent tactic of Muslim users was to defend themselves with comments such as ‘I’m a Muslim and I’m not a terrorist’ or to criticize the ignorance of those tweeting #stopIslam. For example, the 10th most shared Tweet used a Muslim scholar’s quote, ‘If you didn’t study Islam, Please don’t say anything about Islam’ to support their statement ‘Islam doesn’t teach terrorism’ and negate this relationship. This intervention was met by a barrage of memes, illustrated by Figure 12, and accompanied by sarcastic comments such as ‘sure, except for this!’.

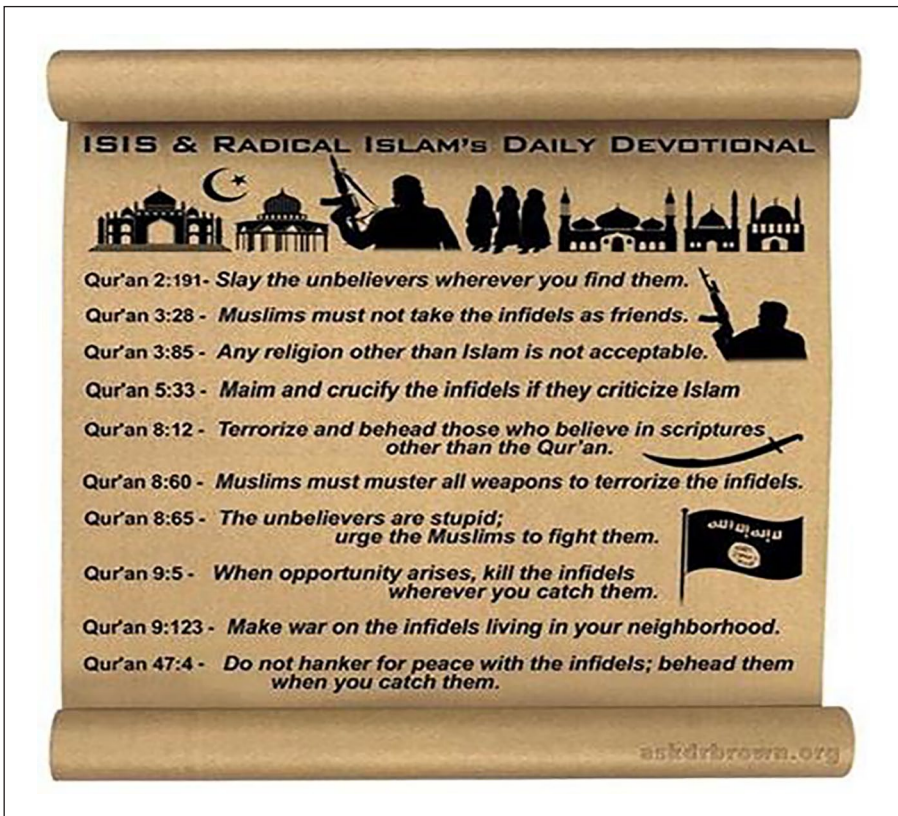


Figure 12. Meme used to assert a relationship between Islam and violence.

The Tweet was shared 1818 times and received 55 negative comments and 11 comments in support of the counter-narrative, with only seven Muslims contributing. The tactic of *support* is evident, but the low proportion of comments by Muslim suggests a lack of engagement with racist hashtags, perhaps in an attempt to reduce its further circulation. This is sometimes explicitly stated, as with recurring calls to ‘stop making this hashtag trend’. However, it is also possible that the affective response here is that of vulnerability in the face of flak from users propagating the original narrative.

It was not universally the case that users refused to engage with attacks on their identity and more emotive responses were also present. There was some evidence of responses that held similarities to tactics employed by would-be allies, but had a more affective dimension due to the way commenters referred to their own identity and knowledge. For example, there were 11 responses by Muslims (of 136 comments) to Figure 13, a meme circulated by a serial contributor to this hashtag who describes using identifiers including ‘Pro-life’, ‘Military’, ‘CCOT’ (Conservative Christians on Twitter), and ‘pro guns’. This meme is accompanied by a statement of defiance against ‘the evil’



Figure 13. Dominant narrative meme.

because the countries shown are ‘all united against Islam’. Tactics by participants identifying as Muslim were to *attack* the poster for his racism, *defend* Islam and Muslims (disassociating them from terrorism), and *dispute* aspects of the imagery, such as the use of certain flags that suggest the allegiance of specific countries (against Islam). These tactics thus resonate with what Abdel-Fadil (2019) describes as ‘affective performances . . . attempts at shifting the balance of power, and reclaiming an object, such as an identity, nation, or religion’.

There was one tweet in this section that did not follow the usual pattern of being silenced by persistent flak. This was posted by the founder of the Quilliam Foundation, a controversial UK-based counter-extremism organization which has been criticized for reinforcing both security policies towards Muslims and the dichotomy between radicals and moderates. Not unsurprisingly therefore, this tweet is both critical of the hashtag itself *and* Islam, suggesting that religious reform is also needed. Probably due to the diversity of people following the Quilliam foundation (other prominent groups such as the Clarion Project³), responses to the tweet are divided and provoke a more nuanced intellectual debate around the problem of reform in Islam. There are some repeat right-wing commenters, but overall this tweet reflects a different audience.

Celebrities and the mainstream media

The support of Muslim participants can also be found in responses to celebrity tweets. As discussed in the methodology, we are defining celebrities here as those whose fame originated in a field other than their use of online media (with our sample including actors, comedians and popular musicians). While recognizing that this distinction between celebrity types has become blurred, it remained useful here for heuristic purposes as – in our sample – microcelebrities tended to derive from the right-wing blogosphere (with their tweets conforming to tactics described in the ‘tactics of the right’ section, above). The dynamics of celebrity uses of the hashtag, in contrast, was distinctive, demonstrating a slightly different trend in the responses compared to other tweets.

Although there were only seven celebrities tweeting the hashtag in the 150 tweets analysed, five of these were in the top 50 most shared due to the number of followers these accounts had accrued. Most of the celebrities tweeted messages supporting Muslims and received mainly positive responses due to the number of fans among their followers, including Muslim fans who thanked the celebrity for their support. There were some similarities between celebrity engagement with the hashtag and other uses (as with many of the regular tweets, the celebrities do not interact with those commenting for instance), but their relationship to the wider counter-narrative was different. As discussed above, the most common trend observed beneath wider counter-narrative tweets was one of ‘trench warfare’ (Karlsen et al., 2017) where those with opposing opinions were brought together, but retrenched existing standpoints rather than engaged in dialogue. Celebrity Twitter use, in contrast, was far more akin to an echo-chamber, pointing to the limitations of such engagements. One example (Figure 14) is the third most shared tweet by a guitarist of a rock band, who describes the hashtag as ‘ridiculous’ and shares the following text:

This highlights the unintelligence of a worrying amount of people. It demonstrates not only astronomical ignorance, but also pertains to an alarming lack of education. Painting all followers of Islam under the same brush of 'terrorism' is as stupid as it is wrong- I presume you don't group all Christians under the umbrella of the KKK? Either way, people who proclaim to be patriotic or doing 'god's work' by condemning entire religions are disgusting. This view reminds me of someone else who shared this mindset... And he ethnically cleansed over six million people. Go worship and preach whatever, or whoever you believe in, but I worship this world's people.

Figure 14. Meme shared by a celebrity: counter-narrative.

This was originally retweeted 3904 times with 6254 likes and 362 comments, most of these agreeing with or thanking the celebrity for his support, aimed at trying to capture his attention. While there were a few interventions from the right, some of these were less hostile (focusing on more general points about political correctness, for example) possibly due to their affiliation to the celebrity (24%). Fan accounts of celebrities, on the other hand, do follow the normal pattern of a backlash against the counter-narrative they tweet, probably due to the wide audience they attract and having less attachment to the author.

As with celebrity Twitter accounts, mainstream media organizations also tended to foreground the counter-narrative against hate speech, but with a slightly different emphasis. When we looked at the hundred most prominent accounts who had shared the hashtag (by numbers of followers), we found that almost a quarter of this list consisted of verified media institutions (22, including *Al Jazeera*, *CNN*, *Nigeria Newsdesk*, *The Independent*, and *The Washington Post*). None of these accounts echoed the hashtag's original meaning in any way (even indirectly), with the majority of accounts instead reporting on the counter-narrative and only one (*Russia Today*) reporting on the existence of the original narrative itself. While two institutions in the top 100 that lacked Twitter's 'blue tick' (including *Breitbart*, which was not yet verified) *did* leverage the hashtag in support of anti-Islamic narratives – portraying it as reflecting a broader public mood that was concerned about 'Islamification' – again, the majority focused on the counter-narrative. The framing of these reports thus suggests it was the counter-narrative's contestation of hate speech, which rendered #stopIslam a newsworthy event.

Unlike other research focused on anti-racist activism (Jackson and Foucault Welles, 2015, 2016), however, our qualitative findings suggest this uptake of the counter-narrative has significant limitations. On the one hand, these media narratives did not feed back into Twitter discourse in a substantive way (with very few of the most retweeted tweets coming from verified news institutions). On the other hand, the way the hashtag was engaged with by the mainstream media did not fundamentally combat long-standing problems of exclusion and discrimination within the press; while contestation of #stopIslam was reported on in a broad sense, this narrative did not offer space for those most affected by Islamophobia to have a platform in the mainstream media (in part due to the demographic of those participating in the counter-narrative, where Muslim voices were outweighed by those of would-be allies). In addition, our qualitative examination of comments beneath the most retweeted tweets points to the danger of optimistically reporting on the contestation of #stopIslam as reflective of a wider public backlash against hate speech or – at least – of the capacity of Twitter to support this backlash. As outlined above, the swift way that positive sentiment was shut down by tightly clustered networks of users points to the danger of uncritically reporting on the ‘success’ of the counter-narrative. Positive framings of responses to the hashtag, moreover, are also in danger of distracting from the culpability of certain strands of the mainstream media for their own role in creating the conditions in which hate speech can become normalized as acceptable public opinion.

Conclusion

Media engagement with counter-narratives against #stopIslam bears out work that has discussed the capacity of activist hashtags (Clark-Parsons, 2019) and anti-racist hashtags in particular (Jackson and Foucault Welles, 2015, 2016) to afford counter-narratives wider news value, here providing an opportunity for alternative constructions of Muslims in the public sphere. Yet the media’s engagement with #stopIslam also foregrounds points of tension; if mainstream media reports were to be believed, although platforms such as Twitter open space for hate speech to circulate, they also offer opportunities for ‘ad hoc publics’ (Bruns and Burgess, 2011) to coalesce in order to *contest* these sentiments. Such narratives, however, run the risk of legitimating social media platforms’ own idealistic self-presentation of themselves as platforms for freedom of speech that (to draw on the language of Twitter’s own terms and conditions) ‘serve the public conversation’ (Twitter, 2019).

Presenting racism and Islamophobia as something that can be ‘debated’ reinforces structural inequalities (Titley, 2019), by appealing to idealized notions of freedom of speech that are predicated on the figure of the unmarked, White, Anglo-Saxon man as being at the centre of Internet use (Nakamura, 2002). The dangers of valorizing ‘debatability’ are underlined by reflecting on responses beneath the tweets of both Muslims and non-Muslims that propagated the original sentiment of #stopIslam. The criticisms of all counter-narrative tweets were relatively similar, presented as criticisms of Islam or of ‘liberalism’ in general, rather than attacking individuals for their religious identity (which also had the consequence that these tweets could avoid contravening hate speech guidelines). As we have elucidated above, however, the *ramifications* of these discourses are

very different in terms of the emotional labour required from Muslim users and would-be allies, respectively.

Depictions of Twitter as contributing to ‘conversation’ are especially dangerous in a context animated by affectively charged (Abdel-Fadil, 2019), antagonistic (Evolvi, 2018) exchanges. In the case of #stopIslam, the reproduction of existing inequalities was compounded by the broader media ecology, specifically the relationship between Twitter and right-wing media: as evident in the emergence of right-wing bloggers as nodes who anchored the tightly bound networks of users circulating hate speech. These dynamics meant Twitter debates ultimately consecrated ‘exclusionary solidarities’ (Wahl-Jorgensen, 2019: 110) via an instrumentalized use of storytelling (Siapera et al., 2015) that did not leave space for dialogue but instead closed it down, by propagating a consistent, negative, narrative about Islam. Our findings, therefore, disrupt idealized Habermasian scenarios that portray the creation of more space for people to speak as the best remedy for hate speech. When reflecting on how to combat hate speech, then, it is important to avoid tactics that inadvertently reinforce this notion of ‘debatability’ in the framing of counter-narratives, perhaps focusing on identifying Islamophobic and racist content *as* hate speech rather than disputing disinformation in the content of tweets in ways that open up avenues for further disinformation to be propagated.

In addition, our findings suggest that even though contesting hate speech can offer scope for reframing mainstream media discourse (albeit briefly), caution needs to be taken when engaging with narratives where the terms have already been set, pointing to the importance of constructing alternative narratives rather than just hijacking existing ones. While, following other work in media studies, we have used ‘counter-narratives’ in its broadest sense in this article, some academics and campaigners (e.g. Blaya, 2019; De Latour et al., 2017) have begun to distinguish between counter-*narratives* (that offer alternative frames) and counter-*speech* (that attacks or denies the original sentiment of a message, but restages it in the process). This narrative/speech distinction is useful in making sense of the limitations of some of the tactical attempts to negate Islamophobic narratives that we identified, which tended to fall into the counter-speech category thus inadvertently recirculated hate speech through the very act of contesting it.

Turning attention to tactics engaged in by those circulating #stopIslam also has broader conceptual implications. While #stopIslam illustrates Twitter’s role as an ‘active agent’ (Kriess and McGregor, 2017) in constituting Islamophobic discourse, it is also important to reflect, conversely, on how actors involved in these debates subtly modulate the affordances of Twitter. To an extent, De Certeau’s (1984) distinction between strategies (the actions of those who define the territory or rules of the game) and tactics (the everyday negotiation of these rules by those inhabiting this space) still holds. This distinction is particularly useful when it comes to understanding how Muslim users and would-be allies tactically negotiated the constraints established by both Twitter and the original narrative. Other tactics, however, depart from work in media theory that has employed the tactics/strategies distinction; they are not the ‘arts of the weak’ that assume a marginal position and can be co-opted by the strategic activities of social media companies (Manovich, 2009). Instead particular tactics are increasingly *intervening* in the affordances of social media platforms in ways that suit the needs of the actors that deploy them.

The significance of right-wing tactics is particularly evident in the way that social media platforms have attempted to combat hate speech and disinformation. While Twitter's monetization of debate means it benefits from controversy to an extent, in order to preserve this flow of capital and avoid censure, content at the heart of these controversies then has to be deleted. These economic relations, combined with the activities of right-wing actors described here, are what have driven strategic changes to hate speech guidelines, led to the deletion of content and resulted in a push for 'techno-fixes' (Marres, 2018) to combat disinformation (such as Twitter's own call for developers to work with them and find algorithmic solutions to monitor the 'healthiness' of communication on the network; Twitter, 2018). These strategies, however, ultimately leave those most affected by hate speech vulnerable to attack by creating 'operational distinctions between organised, extreme racism and ambient, banal everyday race talk' (Siapera, 2019: 2) that can be exploited by right-wing groups to circumvent social media terms and conditions in order to normalize xenophobia (Feshami, 2018). More, these strategies consecrate the position of mainstream media as a 'legitimate' news sources in ways that mask their role in normalizing xenophobia.

The stakes of these developments are high; although the tactical work of a range of different activist groups (including leftist, anti-capitalist, and environmental movements) has been shown to subtly re-shape the affordances of particular media platforms (Feigenbaum, 2014; Poole et al., 2019; Treré, 2019), this study suggests that the dynamics of social media afford more agency to well-organized groups with stronger ties (here the right). In addition, the tactical form of action taken by the right, combined with the rhetorical framing of the content of these interventions, makes these groups appear marginal – fighting against media censorship and the liberal left – at the same time, as these actions are subtly modulating the affordances of social media to consecrate hegemonic values and exclude divergent voices. What is particularly concerning about #stopIslam, therefore, is that it illustrates how the strategies of social media platforms can create conditions that lend themselves not just to the actions, but ideological commitments of right-wing populist groups.

Acknowledgements

We would like to thank Research Assistants Mohammed Al-Janabi and Charis Gerosideris for their assistance with the quantitative analysis and Wallis Seaton for assistance with qualitative analysis.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a British Academy/Leverhulme Trust Small Research Grant (SG161680). Ethical approval was granted by Keele University Ethics Committee, 2016.

ORCID iD

Eva Haifa Giraud  <https://orcid.org/0000-0003-0845-9804>

Notes

1. Gnip is a commercial company specializing in the aggregation of social media content.
2. The intention was to collect data for a month but the funding allowed for some extra time. Following a small pilot study, that showed how counter-narratives dropped off over time, we considered it worthwhile to do this.
3. Its website describes itself as 'A non-profit educational organization providing a platform for human rights activists and for challenging Islamic extremism'; <https://clarionproject.org>. It is based in the United States.

References

- Abdel-Fadil M (2019) The politics of affect: the glue of religious and identity conflicts in social media. *Journal of Religion, Media and Digital Culture* 8(1): 11–34.
- Awan I (2014) Islamophobia on Twitter: a typology of online hate against Muslims on social media. *Policy & Internet* 6: 133–150.
- Baker P, Gabrielatos C and McEnery T (2013) *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.
- Blaya C (2019) Cyberhate: a review and content analysis of intervention strategies. *Aggression and Violent Behavior* 45: 163–172.
- boyd d and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.
- Broersma M and Graham T (2013) Twitter as a news source: how Dutch and British newspapers used tweets in their news coverage, 2007–2011. *Journalism Practice* 7(4): 446–464.
- Bruns A and Burgess J (2011) The use of Twitter hashtags in the formation of Ad hoc publics. In: *Proceedings of the 6th European Consortium for Political Research (ECPR) general conference 2011*, University of Iceland, Reykjavik. Available at: <https://eprints.qut.edu.au/46515/>
- Clark-Parsons R (2019) 'I see you, I believe you, I stand with you': #MeToo and the performance of networked feminist visibility. *Feminist Media Studies*, 1–19. DOI: 10.1080/14680777.2019.1628797.
- Copsey N, Dack J, Littler M, et al. (2013) *Anti-Muslim Hate Crime and the Far Right*. Middlesbrough: Teeside University.
- Cresswell J and Clark V (2007) *Designing and Conducting Mixed Methods Research*. London: SAGE.
- D'Ancona M (2017) *Post-Truth*. London: Random House.
- Dawes S (2017) #JeSuisCharlie, #JeNeSuisPasCharlie and ad hoc publics. In Titley G, Freedman D, Khiabany G, et al. (eds) *After Charlie Hebdo*. London: Zed Books, pp. 180–191.
- De Certeau M (1984) *The Practice of Everyday Life*. Berkeley, CA: University of California Press.
- De Latour A, Perger N, Salag R, et al. (2017) *We Can!: Taking Action Against Hate Speech through Counter and Alternative Narratives*. Revised ed. Strasbourg: Council of Europe.
- De Ridder S (2013) Are digital media institutions shaping youth's intimate stories? *New Media & Society* 17(3): 356–374.
- Deacon D, Murdock G, Pickering M, et al. (2007) *Researching Communications: A Practical Guide to Methods in Media and Cultural Analysis*. London: Bloomsbury.
- Dean J (2009) *Democracy and Other Neoliberal Fantasies*. Durham, NC; New York: Duke University Press.
- Emcke C (2019) *Against Hate*. Cambridge: Polity Press.

- Evolvi G (2018) #Islamexit: inter-group antagonism on Twitter. *Information, Communication & Society* 22(3): 386–401.
- Farkas J, Schou J and Neumayer C (2017) Cloaked Facebook pages: exploring fake Islamist propaganda in social media. *New Media & Society* 20(5): 1850–1867.
- Feigenbaum A (2014) Resistant matters: tents, tear gas and the ‘other media’ of occupy. *Communication and Critical/Cultural Studies* 11(1): 15–24.
- Fenton N (2010) *New Media, Old News: Journalism and Democracy in the Digital Age*. London: SAGE.
- Feshami KA (2018) Voices of white resistance: democratic theory and the task of contending with white nationalist voice. *Open Library of Humanities* 4(2): 391–326.
- Frenda S (2018) The role of sarcasm in hate speech. A multilingual perspective. Available at: https://www.academia.edu/39364015/The_role_of_sarcasm_in_hate_speech._A_multilingual_perspective
- Giraud E (2018) Displacement, ‘failure’ and friction: tactical interventions in the communication ecologies of anti-capitalist food activism. In: Schneider T, Karin E, Catherine D, et al. (eds) *Digital Food Activism*. New York; London: Routledge, pp. 48–68.
- Giraud E (2019) *What Comes after Entanglement? Activism, Anthropocentrism and an Ethics of Exclusion*. Durham, NC: Duke University Press.
- Herman E and Chomsky N (1988) *Manufacturing Consent*. London: Random House.
- Holohan S (2019) Some human’s rights: neocolonial discourses of otherness in the Mediterranean refugee crisis. *Open Library of Humanities* 5(1): 1–27.
- Horsti K (2017) Digital Islamophobia: the Swedish woman as a figure of pure and dangerous whiteness. *New Media & Society* 19(9): 1440–1457.
- Jackson SJ and Foucault Welles B (2015) Hijacking #myNYPD: social media dissent and networked counterpublics. *Journal of Communication* 65(6): 932–952.
- Jackson SJ and Foucault Welles B (2016) # Ferguson is everywhere: initiators in emerging counterpublic networks. *Information, Communication & Society* 19(3): 397–418.
- Karlsen R, Steen-Johnsen K, Wollebæk D, et al. (2017) Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication* 32(3): 257–273.
- Kellner D (2016) *American Nightmare*. Rotterdam: Sense Publishers.
- Kriess D and McGregor S (2017) Technology firms shape political communication: the work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 U.S. presidential cycle. *Political Communication* 35(2): 155–177.
- Lezaun (2018) The public’s two bodies: food activism in digital media. In: Schneider T, Eli K, Dolan C, et al. (eds) *Digital Food Activism*. London: Routledge, pp. 220–227.
- Liao T and Humphreys L (2014) Layar-ed places: using mobile augmented reality to tactically reengage, reproduce, and reappropriate public space. *New Media & Society* 17(9): 1418–1435.
- Manovich L (2009) The practice of everyday (media) life: from mass consumption to mass cultural production? *Critical Inquiry* 35(2): 319–331.
- Marantz A (2018) Inside the daily Stormer’s style guide. *The New Yorker*, 8 January. Available at: <https://www.newyorker.com/magazine/2018/01/15/inside-the-daily-stormers-style-guide>
- Marres N (2018) Why we can’t have our facts back. *Engaging Science, Technology, and Society* 4: 423–443.
- Martin J (2016) The rhetoric of excess. In: Jackson D, Thorsen E and Wring D (eds) *EU Referendum Analysis 2016: Media, Voters and the Campaign: Early Reflections from Leading UK Academics*. Poole: The Centre for the Study of Journalism, Culture and Community, Bournemouth University, p. 21.

- Maxfield M (2015) History retweeting itself: imperial feminist appropriations of 'Bring Back Our Girls'. *Feminist Media Studies* 16(5): 886–900.
- Mendes K, Ringrose J and Keller J (2018) #MeToo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women's Studies* 25(2): 236–246.
- Mercea D, Ianelli L and Loader BD (2016) Protest communication ecologies. *Information, Communication & Society* 19(3): 279–289.
- Modood T (2017) Between nationalism and civilizationism: the European populist moment in comparative perspective. *Ethnic and Racial Studies* 40: 1191–1226.
- Nakamura L (2002) *Cybertypes: Race, Ethnicity and Identity on the Internet*. New York: Routledge.
- Ott BL (2017) The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication* 34(1): 59–68.
- Ouellette L and Banet-Weiser S (2018) Media and the extreme right: editor's introduction. *Communication, Culture & Critique* 11(1): 1–6.
- Poole E (2002) *Reporting Islam: Media Representations of British Muslims*. London: I.B. Tauris.
- Poole E, Giraud E and de Quincey E (2019) Contesting #stopIslam: the dynamics of a counter-narrative against hate speech. *Open Library of Humanities* 5(1): 1–39.
- Rambukkanna N (2015) From #RaceFail to #Ferguson: the digital intimacies of race-activist hashtag publics. *The Fibreculture Journal* 26. Available at: <http://twenty-six.fibreculturejournal.org/fcj-194-from-racefail-to-ferguson-the-digital-intimacies-of-race-activist-hashtag-publics/>
- Ringrose J (2018) Digital feminist pedagogy and post-truth misogyny. *Teaching in Higher Education* 23(5): 647–656.
- Schradie J (2019) *The Revolution That Wasn't: How Digital Activism Favors Conservatives*. Cambridge, MA: Harvard University Press.
- Sharma S (2013) Black Twitter? Racial hashtags, networks and contagion. *New Formations* 78: 46–64.
- Siapera E (2019) Organised and ambient digital racism: multidirectional flows in the Irish digital sphere. *Open Library of Humanities* 5(1): 13.
- Siapera E, Boudouris M, Lenis S, et al. (2018) Refugees and network publics on Twitter: networked framing, affect, and capture. *Social Media and Society* 4(1): 2056305118764437. DOI: 10.1177/2056305118764437.
- Siapera E, Hunt G and Lynn T (2015) #GazaUnderAttack: Twitter, Palestine and diffused war. *Information, Communication & Society* 18(11): 1217–1319.
- Titely G (2019) *Racism and the Media*. London: SAGE.
- Torchin L (2016) Kony 2012 – anatomy of a campaign video and a video campaign. In: Tzioumakis Y and Molloy C (eds) *The Routledge Companion to Cinema and Politics*. London; New York: Routledge, pp. 123–134.
- Townsend L and Wallace C (2016) *Social Media Research: A Guide to Ethics*. Aberdeen: University of Aberdeen. Available at: www.gla.ac.uk/media/media_487729_en.pdf (accessed 23 July 2019).
- Treré E (2019) *Hybrid Media Activism: Ecologies, Imaginaries, Algorithms*. London; New York: Routledge.
- Twitter (2018) Twitter health metrics proposal submission. Available at: https://blog.twitter.com/official/en_us/topics/company/2018/twitter-health-metrics-proposal-submission.html
- Twitter (2019) The Twitter rules. Available at: <https://help.twitter.com/en/rules-and-policies/twitter-rules>
- Wahl-Jorgensen K (2019) *Emotions, Media and Politics*. London: Polity Press.

Author biographies

Elizabeth Poole is professor of Media and Communications at Keele University. Her work has focused on analysing the representational frameworks that dominate in news about Muslims; how these are constructed through political and organisational contexts, and are normalised amongst audiences.

Eva Haifa Giraud is a senior lecturer in Media at Keele University, whose research focuses on frictions and affinities between non-anthropocentric theories and activist practice. Her monograph on these themes, *What Comes After Entanglement?* (Duke University Press), was published in 2019.

Ed de Quincey is a senior lecturer in Computer Science at Keele University. He has researched and taught in the area of online behaviour for 17 years, looking into the usability and impact of websites as well as the uses of the information they collect. He has previously collected and analysed data from social media platforms to investigate disease outbreaks and other health related issues such as hay fever and psychosis.