Informal counter messaging: The potential and perils of the informal counter messaging space

Benjamin J Lee, CREST, Lancaster University UK

b.lee10@lancaster.ac.uk

This work was funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1).

This manuscript has not been published elsewhere, nor has it been submitted simultaneously for publication elsewhere.

Short title: Informal counter messaging

Abstract

Counter messaging has been heavily criticised, not least on the grounds of effectiveness. Whereas current debates are focused on the role of government and large organisations in developing and disseminating counter messages, this paper argues that such approaches overlook the informal production of counter messages. Recognising the appetite for 'natural world' content among those engaged in counter messaging, this paper highlights some of the potential benefits of informal approaches to counter messaging. At the same time, the paper also acknowledges the risks that may result from closer working between counter messaging organisations and informal actors.

Introduction

Counter messaging - communication that seeks to disrupt the messages disseminated by extremist groups and individuals - has become a core component of counter terrorism strategies. Thus far, formal programs, often directed by governments or civil society organisations, have dominated the discussion of counter messaging. This paper seeks to expand this discussion by highlighting the potential for informally produced counter messages to contribute to wider policy goals. The paper concentrates on some of the potential benefits and risks of counter messaging by informal actors. Informal content has the potential to be seen as more credible by audiences, as well as to be more aggressive in its messages, for example by isolating and ridiculing individual extremists and organisations. This comes at a cost, as informal content creates increased risk of backlash against creators, and of encouraging hate speech.

Theoretically, this paper builds on insights from nodal governance theory that see security as being increasingly provided by networks that include public, private, and citizen actors.[1] This paper is also a direct response to the move by some formal organisations engaged in counter messaging to use informally produced content rather than producing their own in-house material.[2] This paper does not suggest that informally produced counter messages are superior to formally produced ones, or that informal messages can be relied upon as a substitute. However, any comprehensive understanding of counter messaging is incomplete without acknowledging the potential of content produced by informal creators, and the potential damage that may come from co-opting it into a formal campaign. Although governments have taken a hand in both directly producing counter messages, and

mobilising wider networks of influencers, these attempts have drawn criticism from both the press and political activists.[3]

What is counter messaging?

Counter messages seek to undermine the messages presented by extremist groups. Existing literature on the subject reveals a broad range of understandings and different terms in use. Peter Neumann, writing on potential responses to online radicalisation in the US, identified counter messages as a potential tactic to reduce demand for violent extremist material.

> *"Broadly speaking, counter-messaging may involve challenges to the violent extremists' ideology and to their political and/or religious claims; messages that aim to "mock, ridicule or somehow undermine their credibility"; contrasts between violent extremists' grandiose claims and the reality and/or consequences of their actions; or positive alternatives that cancel out or negate the violent extremists' ideology or lifestyle."*[4]

Within this definition there are multiple types of message: disputing claims, undermining extremist groups, contrasting rhetoric and reality, and promoting positive alternative ideologies. Other approaches have concentrated more heavily on media and the delivery of messages, including making specific references to social media, as well as religious authority. Setting out elements of a counter messaging strategy designed for use against ISIS, Pelletier et al describes some core components:

> *"Defeating ISIS will be a multi-faceted long-term effort to delegitimize the movement and undermine its radical interpretation of Islamic Law, while inhibiting their ability to persuade and inspire followers. As with messaging, effective counter-messaging is as much a technique as it is a process, and a counter-messaging (CM) strategy also involves a combination of written / oral communications, reinforced by actions and behaviors and then propagated through the effective use of social media."*[5]

Other approaches have varied the terminology. In particular, narrative and counter-narrative has been the main descriptor used in many accounts. Narrative refers to a set of events in sequence, and is more expansive than message, which can refer to isolated claims.[6] Ferguson differentiates between counter narratives as a term deployed in the 'CVE literature' and 'alternative' strategies focusing on existing media and journalism, including drama.[7] Even beyond message and narrative, there are a plurality of descriptions for activities, one report suggests the concept of 'alter-messaging' focusing on 'alternative content to the ideology of terrorism'.[8] In this article, the use of the term counter message is intended to enable a broad analysis. Subsequent discussion of informal counter messages will demonstrate how diverse content can be. Little of the content analysed for this paper would fit the description of a narrative, with actors instead often focusing on specific issues as opposed to longer stories.

One common thread in discussions of counter messaging is a lack of specificity in describing audiences. Leuprecht et al, for example, suggest differentiating between individuals who are closer to the top of the 'pyramid' (of radicalisation) or are likely to be so, and those further down who are seen as potentially at-risk.[9] Neumann talks

about audiences that are 'potentially vulnerable to becoming radicalised'. [10] There is little evidence of consideration of the impact of counter messages beyond audiences who may potentially be vulnerable to radicalisation, despite the impact of extremist messaging, most notably terrorism, on both potential victims and wider society.[11]

Counter messaging fits broadly under the heading of countering violent extremism, a broad policy area that encompasses a range of activities. Harris-Hogan et al aim to use public health models to frame CVE for policy makers, identifying primary, secondary and tertiary interventions.[12] Tertiary interventions seek to enable disengagement from violent extremist networks, for example through deradicalization and disengagement programs. [13] Secondary interventions aim to intervene where individuals are displaying 'symptoms' of radicalisation.[14] Given that individuals targeted by these interventions are unlikely to have committed any offence secondary interventions are prone to being the most controversial. Primary interventions are preventative and aim through training and education to reduce the prevalence of violent extremism.[15] Counter messaging can be seen as a form of both primary and secondary CVE depending on intended audiences. In many cases it is aimed at both those already engaged with violent extremism and those at risk of engagement.[16] There is also further differentiation over the goals of counter messaging, varying between changing minds and changing behaviour.[17]

In practice, counter messaging campaigns can manifest in many different forms, but analysis tends to be confined to campaigns that have some kind of formalised support. Examples include Abdullah-X, a cartoon avatar that features in online 'comics', as well as in a series of 22 YouTube videos. Titles include *The Real Meaning of Jihad*, and *Freedom of Speech vs Responsibility*. The site and videos are well-

presented, and feature references to media engagement by the content creator, reportedly a former extremist. The project is listed as 'made possible by jigsaw', an incubator within Google's parent company, Alphabet. [18] No videos have been produced by since October 2016. The Global Survivors Network focuses on testimony from victims of terrorism. The network produced the documentary film *Killing in the Name*, which was nominated for an Oscar in 2011. The network also had a presence on YouTube, Facebook, and a website (which is now defunct). The Facebook page has not been updated since November 2015. A case study by the Institute for Strategic Dialogue describes the project as 'seeded' following a September 2008 UN symposium, with the aim of sharing victim testimonies to vulnerable communities.[19] Both of these campaigns now appear defunct, although the content they have produced persists online.[20]

Critiques of counter messaging

There have been three main related critiques of counter messaging: strategic effectiveness, normative, and capability-based.

Much of the criticism of counter messaging has stemmed from a perceived lack of strategic effectiveness. This can be framed as questioning the role of messaging in general. The basis of this critique is that policy-makers have over-estimated the importance of propaganda in driving violent extremism, and therefore assume that the solution is to promote the alternate view. [21] Glazzard, in a paper for the International Counter Terrorism Centre, excoriates 'counter-narrative theory', noting the extent to which government, think tanks and advocacy organisations are organisationally committed to the idea and the limited evidence base.[22] Other

critiques of effectiveness dwell on the scale of the task, suggesting that the volume of extremist material makes developing counter messages a drop in the ocean.[23]

These criticisms are further compounded by the difficulties in measuring the effectiveness of counter messaging campaigns. Although raw metrics are often used to support arguments for effectiveness, knowing the final impact of exposure to counter messages is very difficult, particularly in a real-world setting.[24] Not least because, depending on the intended targets of counter messaging campaigns, effective outcomes can be based on non-events. In the preventative space for example, this could mean dissuading an individual from engaging with an extremist organisation at all.

These are valid criticisms. However, much depends on what counter messages are expected to achieve. Involvement in extremist organisations is dynamic and granular. At any one point there are multiple audiences inside and outside an extremist milieu. Some may be committed activists, others may be wavering, others still contemplating deeper engagement. There may also be potential members who are yet to even hear of a group, a largely indifferent public, fearful potential victims, and ideological opponents that may even be involved in their own forms of extremism. Despite an assumed ability to micro-target content, counter messaging campaigns are not limited to a single audience, and leakage from campaigns has the potential to impact unintended audiences. This potential only becomes greater when material is published online where it can be re-posted and remixed to different effects. The effects of counter messaging on these audiences are not well understood, but there at least needs to be a recognition that despite the stated aims of counter messaging campaigns, that they will impact on a variety of audiences. In addition to

these considerations, we might also consider the alternative of remaining silent, and the ramifications of allowing extremist messages to go (formally at least) uncontested.

As well as questioning the effectiveness of counter messaging campaigns to date, there are deeper questions on the normative aspects government involvement. Although few would question the role of the government in confronting terrorism, the indistinct margins between radical milieus and support for violence raise awkward questions around the extent to which interventions are warranted.[25] Richards, for example, argues for a distinction between "extremism of thought' and 'extremism' of method'.[26] Conflating non-violent ideology with violent methods is problematic from a policy angle, but equally distinguishing between non-violent and violent ideology, and crucially the trajectories between the two, is also extremely difficult. These nuanced and at times invisible distinctions create extremely muddy waters for governments looking to either counter extremist ideologies directly, or do so through proxies. Counter messaging sits awkwardly on the dividing line between legitimate counter terrorism and publicly unacceptable ideological engineering. More recent shifts in UK government policy, including the 2015 UK Counter Extremism Strategy, seem to indicate an increasing focus on extremism as opposed to violent extremism.[27] Although this move has been criticised by the Joint Committee on Human Rights for an assumed 'escalator' between extremism and terrorism, and the definition of extremism.[28]

Counter messaging runs the risk of being equated with propaganda. Propaganda has historically been seen in democratic states as the preserve of undemocratic opponents.[29] In the aftermath of World War One, the term propaganda became

synonymous with dishonesty.[30] In some accounts propaganda is used in democratic states either for dubious ends,[31] or as part of everyday communication.[32] Some make the case that promotional culture is now pervasive, and consequently the term propaganda has served its usefulness.[33] Despite the normalisation of propaganda, recent incidents have also showcased public alarm at both CVE policy and counter messaging. In the UK, Prevent, a CVE strategy, has already been much maligned both by the press and civil society organisations on the grounds that it unfairly targets Muslims, causes alienation and is ultimately counter-productive.[34] Focusing specifically on counter messaging, the Home Office-based Research Information and Communication Unit (RICU) was heavily criticised for their *Help for Syria* campaign which distributed leaflets and targeted student events without identifying itself as being government-backed.[35] Clearly in this case, it was felt that acknowledging government involvement would immediately sour audiences on the message, but the revelation may have done further damage to the UK Government's credibility, suggesting the campaign may even have been counter-productive. Examples of RICU's work have been seized upon by anti-Prevent campaign organisations, including the highly partisan pressure group CAGE, which in 2016 published a report *We Are Completely Independent*, in which claims of covert government support were made against several campaigns. The overall tone of the report is highly critical.[36] The risks of 'sock-puppet' organisations, i.e. organisations claiming to be independent but working to advance specific agendas, have been identified as including worsening trust in government over time, as well as undermining the credibility of other organisations without official ties.[37]

A final criticism stems from the lingering doubts about the capabilities of counter messaging actors to produce material that will resonate with audiences. One account

of counter messaging in the US identifies a 'preachy' tone as being a key problem with audiences.[38] Meanwhile, governments are seen by many as having a 'credibility gap' with target audiences, and efforts to engage may be dismissed out of hand, or even potentially entrench extremist beliefs.[39] Writing on countering ideological support for terrorism, Herd and Aldis argue:

> *"This* [countering ideological support for terrorism] *is not an appropriate task for governments to undertake, but on the evidence of our case studies appears best carried out by indigenous religious or other civil society organizations. Too obvious governmental efforts in this field, too close cooperation with moderate religious associations within a region or state, will only serve to delegitimize them in the eyes of the population."*[40]

Civil society organisations are not tainted with the same lack of credibility that governments are, however, civil society organisations are very often aligned with governments, seeking to obtain favour and resources. The perception (often ideologically skewed) is potentially one of a counter extremism industry, composed of charities, think-tanks, and other organisations, all seeking to engage with high-priority counter-terror efforts. Schmid suggests that many civil society organisations are dependent on government for funding, and are staffed by employees who move between governmental and non-governmental organisations.[41] Likewise, Tierney argues that, in the North American context, governments are still seen as the 'ultimate drivers' of CVE efforts.[42] To illustrate this point, the Abdullah X campaign (described above) was described in one (highly ideological) blog as follows:

> *'With further research one realises that the connections of the "Abdullah-X" project reaches into the global counter-radicalisation industry which promotes neoconservative, Zionist aims.'*[43]

Of further concern should be the cultural background of those working in civil society organisations. Although the prime concern of counter messaging is ideological, this is likely heavily embedded within cultural practices particular to target groups. These can include language, customs, music and art.[44] The question remains: how much credibility do government, and government-backed approaches, to CVE and counter messaging have for audiences?

Formal and informal counter messaging

It is fair to say that analysis of counter messaging, along with CVE more broadly, thus far has been focused almost entirely on organised programmes, often those with financial connections to government and larger NGOs. The presence of government funding in this space has led to a focus on discrete programs undertaken by identifiable organisations with measurable outcomes.[45] The need to justify the use of public money and measure success funnels funding to programmes that can produce measurable results.[46] This focus on formal programmes risks ignoring some of the most potentially useful contributions to counter messaging.

In contrast to the current academic and policy fixation, much counter messaging work is done informally by citizens with no connection with government security policy or any wider community organisations. At the micro-level this means

conversations with friends and family, discussions around the dinner table, in clubs, community centres, and in the back-rooms of pubs. Research based in Indonesia has highlighted the informal role of women in challenging extremism, suggesting that they constitute an important resource that has been allowed to go 'under the radar'.[47] Although these micro counter messages remain difficult to access, the ease and availability of digital communication platforms has resulted in a corpus of readily accessible counter messages in formats including video and social media accounts. In effect, alongside government and NGO efforts at counter messaging, there is an informal sector of actors producing digital content that is critical of extremist messages. Based on the existing criticism of government-aligned counter messaging, there is good reason to believe that informal counter messages are likely to differ significantly from more formalised approaches to counter messaging.

Probably the most infamous example of informal messaging in the UK emerged following a 2011 demonstration by the counter jihadist street group the English Defence League. At the event, Press TV interviewed a (possibly drunk) EDL supporter during the demonstration asking why he was protesting. His reply was garbled and slurred. One reference, possibly to 'Islamic rape gangs', was heard as 'muslamic ray guns', and this became the hook of a music video created by auto-tuning the original interview.[48] The resulting video, produced by Alex Vegas, attracted over 1.9 million views. The song, and the phrase became a running joke at the expense of the EDL and is available to buy as a t-shirt. Defending himself in the comment section of the original video, creator Alex Vegas said: "All I did was make him sing, I didn't change what he said".

While Muslamic Ray Guns serves as perhaps the best known, there are additional examples of counter messaging which have received less attention. The YouTube channel Veedu Vids was established in 2015 with the motto 'let's beat bigotry with a smile'. At the time of writing the channel has over 2,300 subscribers, and features 30 videos ranging from one minute 30 seconds, to 6 minutes and 45 seconds long. The videos deal primarily with topics around Islam and Islamist extremist narratives and speakers, but also features some videos focusing on 'alt-light' figure Milo Yiannopoulous, as well as controversial evolutionary biologist Richard Dawkins. Titles include *ISIS Appeal: Don't mock ISIS*; *Milo Yiannopoulos VS Anjem Choudary: Is Islam compatible with the West?*; and *Zakir Naik: Are Nursery Rhymes Halal? (PARODY)*. Videos tend to focus on parody and imitation with the protagonist (presumably the channel owner) taking on all the roles in short sketches. An early video – *Abu Haleema Trailer (PARODY)* centres on a montage of shots set to music (Carmina Burana) in which the protagonist seemingly imitates the London-based militant Islamist Abu Haleema. The parody of Haleema featured his trademark beard and hand gestures. Haleema rose to some prominence following a Channel Four Documentary – *The Jihadis Next Door* – which followed Haleema through various legal troubles and gave some insight into the production of YouTube videos. Despite the media exposure, Abu Haleema's now seemingly deleted YouTube account had around 1,600 subscribers (in early 2017).

In addition to video, social media accounts have also been a source of informal counter messaging. In some cases these can take the form of parody accounts such as 'Britain Furst', a successful Facebook page parodying the far-right group Britain First.[49] The page specialised in developing alternative versions of the Facebook memes that helped to popularise Britain First on social media. Other social media

presences are more straightforward in their opposition such as Exposing Britain First, which has accounts on multiple social networks, and the Facebook group Muslims against Daesh. One important caveat on these examples is the tendency of government and society actors to conceal the source of counter messaging campaigns. Given these practices, it is not possible to say for certain that any of these examples constitutes genuine counter messaging and is not a component of a broader government or civil society initiative.

The potential for formal and informal collaboration

Despite the inherent murkiness on the communications environment, at this juncture there is little evidence of coordination between formal and informal counter messaging. However, the informal sector has been identified as a potential source of content that could be co-opted by formal organisations. A 2016 report by the Institute for Strategic Dialogue, a prominent counter extremism organisation, recommended that 'natural world' content could be used to get around the bottlenecks presented by the need to create high-quality original content.[50] A later report on the extreme-right also argued for greater counter-culture specific knowledge:

> *'They* [counter speech measures] *must penetrate alternative platforms and burst extreme-right bubbles with campaigns that build on a thorough understanding of internet culture and counter-cultures.'*[51]

Likewise, the Jigsaw-backed 'Re-direct method' advises against the creation of original content in service of new counter messaging campaigns:

> *'Campaigns to confront online extremism don't necessitate new content creation. The best part of the research beyond identifying ISIS's recruiting narratives and the content categories most likely to debunk them was that it surfaced hundreds of online videos in English and Arabic that were already uploaded to YouTube, and that would not be be* [sic] *rejected outright by our target audience.'*[52]

There is also a strong theoretical case for building closer ties between formal and informal content creators. Closer cooperation between security practitioners and private citizens has been identified as a possible model of security provision. In a cyber-security context, nodal governance models have been applied as tools for analysing the actions of citizens who have used technology to take on the role of criminal investigators.[53] The concept has also been directly applied to a citizen-led terrorism investigation. The investigation into the 2013 Boston marathon bombing initiated by a number of 'digilantes' using the social media site Reddit provided examples of ad-hoc organisation, deployment of specialist resources to analyse the devices used, and crowdsourced investigations into the identities of suspects.[54] Ultimately, the Reddit-based investigation earned some interest from the FBI, but identified the wrong suspect. Nevertheless, the Reddit affair stands as an example not only of the public's willingness to use technology to intervene in security matters, in this case an investigation, but also the extent to which authorities are powerless to stop them.

> *'…the proverbial genie is out of the bottle: the Internet has created an environment in which the public can and will choose to play a role in public criminal and other investigations that capture its interest.'*[55]

Nodal security frameworks identify individual actors or groups – nodes – which then form networks, each node able to apply its own capital to a collective problem.[56] Capital can include technology, political and social relationships, and resources.[57] In the case of the investigation into the Boston Marathon Bombing, Reddit users were observed to contribute specialist knowledge to the investigation, as well as to pick through media (e.g. photographs) associated with the attacks.[58] Applying this framework to counter messaging, informal content creators bring both symbolic capital, in the form of legitimacy, social capital through their relationships with audiences, and cultural capital stemming from the content they produce. When combined with the critiques of existing formal counter messaging the benefits of informally produced counter messages, in the form of credibility and content, become apparent.

Credibility

The leading critique of formalised counter messaging has been credibility. This applies both to direct communication by government, and to indirect attempts to gain influence through allied groups.[59] Literature on online persuasive communication emphasises the importance of source credibility in persuasiveness.[60] Analysing the evidence on persuasive communication online, Wathen & Burkell identified 26 factors under five headings: source, receiver, message, medium, and context, that could affect persuasion.[61] They further argued that credibility was a

multi-dimensional concept, composed primarily of perceptions of a source's "trustworthiness" and "expertise", but also including the source's presentation (dynamism, likeability and goodwill).

Source credibility factors can also interact with aspects of the message itself. Where persuasive messages differ extensively from already held beliefs, the persuasive effect of messages can be reduced.[62] This confirms more theoretical accounts which have long held that successful persuasive attempts are ones that work with established attitudes, rather than directly contradicting them.[63] Persuasive materials that were too firmly rooted in the interests of the message originator, and not the audience, were unlikely to find favour. Analysis of US attempts at countering misinformation has raised the problem of a 'the credibility deficit opened up by the tension between rhetoric and practice in the US War of Ideas'.[64] For many audiences, supporting, being insufficiently critical of, or failing to address government policy, is likely to be a turn off.

In contrast, audiences may be making different judgements about source credibility in response to content from informal actors than formally aligned ones. Whereas formal counter messaging risks being labelled as biased, or out of touch, informal actors can present themselves as free from bias (at least bias towards the government line), and more in touch with their audiences. In part, some of this credibility may come from accepting assumptions that are incompatible with government support, for example conservative religious positions, or attitudes towards foreign policy. Given the scope and scale of the potential counter messaging space, the most credible spokespeople are unlikely to align themselves with formal counter messaging programmes. The ones that do may risk longer term damage to their credibility.

Content

Schmid has suggested that being one-step removed from government allows civil society organisations greater leeway in terms of the content that they produce.[65] In the informal counter messaging space there are no checks and no standards. Counter messages produced by non-aligned actors are free to include personal attacks on specific individuals, use humour, and even sympathise explicitly with extreme political views. One potential strategy is to discredit violent extremist organisations, or to target individual extremists for ridicule.[66] This strategy may be difficult for larger organisations to engage with without being perceived as bullying, and risking adding legitimacy to extreme actors by acknowledging them and their claims. Some researchers have called for a more robust approach to content creation to match the content produced by extreme groups i.e. 'to fight fire with fire'.[67]

As well as the freedom to produce coarser content, informal counter messaging actors are also free to experiment with different forms of content. While internet memes have come to mean something different from the original use of the term meme, the idea of harnessing the power of seemingly ephemeral trends to sway the public mind has been advocated by several campaigners.[68] Memes, according to Shifman, are characterised by three factors: they spread through interpersonal contact but influence wider culture, they reproduce through imitation but not through exact replication, and they compete for survival with one another.[69] In the past, the concept of meme warfare has also found favour with anti-corporate groups. Kalle Lasn used the term 'meme wars' and 'meme warriors' to set out his vision of a second, anti-corporate, American revolution. Mainstream political parties have copied the form, if not the underlying philosophy of memes in their production of

online political posters designed to be shared on social media platforms such as Facebook.[70]

Meme-inspired approaches may also have value for counter messaging, Laura Huey analysed the production of messages on the social media platform Twitter during the 2015 kidnap of two Japanese citizens by ISIS. Huey highlights the satirical nature of 'political jams' by focusing on the trend of responding to the videoed ransom demand by reconfiguring images to show the kidnappers and the hostages in various absurd situations: serving sushi, carving kebab meat, and reversing the position of kidnappers and hostages.[71] In particular, Huey identifies the role of these images as allowing audiences to respond to fear with humour. Muslamic ray guns has also arguably attained the status of a meme.

In contrast. where government backed communication has attempted to replicate this style of communication it has been less successful. A video produced by the Centre for Strategic Counterterrorism Communications, in the US State Department called Run, Don't Walk to ISIS land, featured footage produced by ISIS itself, including executions. The tone of the video was sarcastic, intended to highlight the hypocrisy of ISIS's behaviour, in particular its treatment of Muslims. The video provoked a public backlash, criticised on the successful HBO show *Last Week Tonight* as 'ironic propaganda'.[72] The ensuing criticism caused the Centre to change direction, limiting themselves to a more fact-based approach to counter messaging.[73]

Collaboration and risk

On paper, there is much to gain from greater collaboration between formal and informal content creators. However, it is not clear what form any formalised support

for informal counter messaging actors would take. It may be possible that strategies such as the re-direct method intend to co-opt content produced by others, promoting it through paid advertising, without building any more concrete relationship with creators. Alternative models could theoretically range from building loose relationships, provision of tools and training, to incorporating content creators into existing organisations. However, any closer working would entail risk. These risks can be further divided into four types: personal risks, risks to targets, strategic risks, and reputational risks.

Firstly, informal actors are often getting down and rolling in the mud with extreme political actors and groups. In highly localised contexts they may risk their own safety, for example one video in which the protagonist confronts Abu Haleema and challenges his messages, the protagonist is clearly concerned for his safety. The protagonist is audibly nervous, and the video is filmed entirely from the protagonist's point of view whilst concealing his identity.[74] The channel the video is posted on contains only that one video and there is no further way to contact the author. In addition to physical security, non-aligned actors also open themselves up to responses from extreme actors online. Veedu Vidz composed a video in which he responded to negative comments received online. Comments included: 'Son of a bitch, everyone start reporting this arsehole production'.

A second risk is the risk to specific targets of content. Parody videos of Abu Haleema for example stimulated dehumanising language in comments sections, and calls for his death. This risk can be, at least in part, be viewed as being a component of reciprocal radicalisation or cumulative extremism. This is the hypothesised interaction between extreme political and religious groups, in which different forms

of extremism fuel each other.[75] Content designed to ridicule extreme political and religious positions, in particular content that singles out individuals, may raise tensions within communities and contribute to escalations in the form of either rhetoric or violence.[76] Analysis of counter messages posted on the social networking site Facebook, for example, found a high proportion of 'non-constructive counter speech' on some pages.[77] The trend for some extreme movements to present themselves as moderate opposition groups to other extreme movements is another good illustration of this problem.[78] The dividing line between counter message, and in some cases hate speech, is heavily trafficked, and the difference between counter speech and provocation is not always immediately apparent.

The third risk plays out on a broader strategic level. Isolation and distance from wider society has been seen as a potential factor in the move towards violence among some groups. Everton's analysis of the Hamburg cell for example uses the concept of socio-cultural tension as a factor in network closure: the hiving off of extreme clusters of actors in networks into their own echo chambers.[79] Counter messaging risks exacerbating and increasing the sense of isolation for those already engaging with extreme groups. While the intention to 'de-cool' extreme groups may be viable for those who have not yet engaged, for others that already identify with extreme groups, or who are already actively involved, then ridicule will likely do little to persuade them to desist. This contrast is a good example of the complexities of CVE. In this instance one form of CVE – primary counter messaging – could well negatively impact secondary and tertiary CVE – attempts to either persuade extreme actors to disengage, or to rehabilitate those who have already engaged in violence.

Finally, the value of informal counter messaging content is precisely its perceived distance from a broader policy agenda. Closer working relationships raise similar risks to closer working between government and civil society: that informal creators will be branded as cogs in the wider CVE machine by audiences. While the investigatory capacity of citizen actors in the Reddit affair was relatively stable even during their involvement with the authorities, the source credibility enjoyed by informal counter messaging actors is much more likely to be damaged as a result of collaboration, or even unwitting co-option, by more 'establishment' agencies. There is the risk that harnessing counter messaging produced by informal actors will risk damaging the effectiveness of content itself.

Reputational risk also goes both ways. While government-backed counter messages need to serve government policy, informal actors have unknown motivations, and may even be actively hostile to wider CVE policy agendas. In some cases, content may not be produced for explicitly political ends, but instead it may emerge from a range or mixture of motivations, including commercial gain and entertainment. *Father Daughter Ad*, produced by long running US comedy institution Saturday Night Live, is arguably an example of counter messaging content produced for commercial ends.[80] In other cases those creating counter messaging content may be closer to agents of chaos. Interviews with an actor on a social media account mocking the far-right, reveals that they were motivated by 'shits and giggs' rather than by any political conviction.[81] Even where content is produced for political ends, this does not mean that the ideological outlook of activists producing the content is compatible with collaboration with state-backed agencies. Consider for example anti-fascist movements which can be simultaneously opposed to far-right extremism, but equally committed to opposition to unjust state practices. For example, Unite Against

Fascism has been critical of the Prevent strategy which, it argues, 'places an eye of suspicion' on Muslim communities.[82]

Conclusions

There will always be ambiguity surrounding the effectiveness of counter messaging as a tool for challenging violent extremism. Primarily, the challenges of counter messages come from the indistinct audiences. Although some level of targeting is possible, there is no way to know if exposure to counter messages serves to reduce future involvement in violent extremism. Despite this, challenging the narratives of violent extremists remains a policy goal. In the UK, the government has invested significant resources into supporting counter messaging both by large civil society actors as well as from smaller community groups.

The analysis here is inevitably focused only on mediated communication because it is accessible to a researcher outside of the space. This analysis does not include the work being undertaken beneath this level in families and communities. The primary argument of this paper is that, in addition to this acknowledged level of counter messaging, there exists an informal level of non-professionals creating counter messaging content. In some cases actors are likely just speaking their minds and have little idea of how the content they produce may align (or not) with broader policy goals. What's more, given the proximity of these creators to potential audiences, and the lack of restrictions on the content they produce, there is the possibility that informal counter messaging content may differ from government-supported content. This paper has suggested that, in some cases, the independence of informal actors compared to government aligned ones may boost their credibility. Not only are informal actors more independent, but they are likely to be seen as more

authentic by audiences. Key examples have also raised the possibility of content based around memes, humour, and personal attacks that may become harder the more bureaucratised content creation becomes. Equally, official content is ultimately tied to policy goals and societal norms that many members of target audiences, not just violent extremists, reject.

The price for this edgier and potentially more influential content, is risk. Non-aligned counter messaging creates risks for both actors and their targets. Ridiculing an extreme figure or group may potentially inflame tensions and risks giving support to opposed extreme views, even where authors do not intend it. Likewise, 'de-cooling' extreme groups may work for those in the audience that are yet to engage, but it may risk further alienating those already involved. Research on group radicalisation suggests that isolation from wider social networks may be a factor in the move towards violence. For some audiences, informal counter messaging may be counterproductive.

There are good theoretical and practical reasons to expect that officially aligned organisations may move to co-opt at least some informal content. Citizen actors, supported by communications technology, are free to bring their own resources to bear on problems, and state agents are likewise able to make use of these. The nodal governance argument is, however, heavily dependent on actors sharing goals which may not always be the case for counter messaging actors. Also, the resources of informal counter messaging actors, specifically their credibility and their free-wheeling content, may not be appreciated or survive in organisations that need to account for themselves publicly.

Finally, it is worth considering the policy ramifications of this research area. Although it is tempting to speculate about the potential for informal content to inject fresh life into officially recognised counter messaging efforts, more research is needed to better understand the experience of informal counter messaging actors, whatever they may call themselves. Specifically of interest are their motivations and how they see themselves in relation to questions of extremism, terrorism, and policy. Although there is clearly interest from some groups in co-opting 'natural' content, no one has asked how content producers would feel about this. It's also not clear what level of support can be offered if the benefits of informal activism are to be preserved. At what point will informal content become similarly inauthentic? These risks need to be understood and managed if closer working is to serve wider policy aims.

Perhaps a more important point however is to remind ourselves of the inevitability of counter messaging. Extreme groups are by definition in the minority. Societies will always react negatively to extreme messages, and given the tools to express their opinions some citizens will do so. Furthermore, it is likely that counter messaging content is regulated to some extent by the size of perceived threats. Where extreme content goes viral and breaks into mainstream networks, then members of those networks will respond. Where extreme content remains confined to obscure networks of supporters, then there will be fewer critical responses created. Informal counter messaging exists in equilibrium with extremist messaging, regardless of government policy.

Notes:

---

[1] Johnny Nhan, Laura Huey and Ryan Broll, "Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings," *British Journal of Criminology* 57(2) (2015), pp. 341-361; Laura Huey, Johnny Nhan, Ryan Broll, 'Uppity Civilians' and 'Cyber-Vigilantes': The Role of the General Public in Policing Cyber-Crime," *Criminology & Criminal Justice* 13(1) (2012), pp. 81-97; Benoît Dupont, "Security in the Age of Networks". *Policing and Society* 14(1) (2004), pp. 76–91.

[2] Tanya Silverman, Christopher J. Stewart, Zahed Amanullah, Jonathan Birdwell, *The Impact of Counter-Narratives: Insights From a Year-Long Cross-Platform Pilot Study of Counter-Narrative Curation, Targeting, Evaluation and Impact*. London: Institute for Strategic Dialogue. (2016). Moonshot CVE, Quantum Communications, Valens Global & Nadia Oweidat, Jigsaw, *The Redirect Method: A Blueprint for Bypassing Extremism* (n.d.).

[3] Ben Hayes, and Asim Qureshi, "We Are Completely Independent: The Home Office, Breakthrough Media and the PREVENT Counter Narrative Industry", (CAGE: London, 2016).

[4] Peter Neumann, "Options and Strategies for Countering Online Radicalization in the United States." *Studies in Conflict & Terrorism*, *36*(6) (2013), p447. Neumann draws on a working paper from the Institute for Strategic Dialogue for the wording here.

[5] Ian R. Pelletier, Leif Lundmark, Rachel Gardner, Gina Scott Ligon & Ramazan Kilinc, "Why ISIS' Message Resonates: Leveraging Islam, Socio-Political Catalysts and Adaptive Messaging," *Studies in Conflict & Terrorism,* 39(10) (2016), p891

[6] Andrew Glazzard, *Losing the Plot: Narrative, Counter-Narrative and Violent Extremism,* (International Centre for Counter-Terrorism, The Hague, 2017).

[7] Kate Ferguson, *Countering Violent Extremism Through Media and Communication Strategies*, (University of East Anglia: Partnership for Conflict, Crime & Security Research, 2016).

[8] Michael Pizzuto, *Alter-Messaging: The Credible, Sustainable Counterterrorism Strategy*, (Goshen, IN: Centre on Global Counterterrorism Cooperation, 2013).

[9] Christian Leuprecht, Todd Hataley, Sophia Moskalenko & Clark Mccauley 'Containing the Narrative: Strategy and Tactics in Countering the Storyline of Global Jihad' Journal of Policing, *Intelligence and Counter Terrorism*, 5(1) (2010), pp.42-57.

[10] Neumann, "Options and Strategies for Countering Online Radicalization in the United States,"

[11] Alexander Schmid & Janny De Graaf. *Violence as communication: Insurgent terrorism and the Western news media,* (Beverly Hills: Sage, 1982).

[12]Shandon Harris-Hogan, Kate Barrelle, and Andrew Zammit, "What is countering violent extremism? Exploring CVE policy and practice in Australia," *Behavioral Sciences of Terrorism and Political Aggression*, 8(1) (2016), pp. 6–24.

[13] Daniel Koehler, *Understanding De-Radicalization* (Oxford, Routledge, 2016)

[14] Harris-Hogan, Barrelle, & Zammit, "What is countering violent extremism? Exploring CVE policy and practice in Australia".

[15] Ibid.

[16] Omar Ashour, "Online De-Radicalization? Countering Violent Extremist Narratives: Message, Messenger and Media Strategy," *Perspectives on Terrorism*, 4(6) (2010), pp. 15–19.

[17] Rachel Briggs, Sebastian Feve, *Review of Programs to Counter Narratives of Violent Extremism* (London: Institute for Strategic Dialogue, 2013); Pizzuto, *Alter-Messaging: The Credible, Sustainable Counterterrorism Strategy*.

[18] Jigsaw, "Disrupt online radicalization and propaganda" (n.d.) https://jigsaw.google.com/projects/#abdullah-x accessed 4/10/17

[19] Institute for Strategic Dialogue, "Global Survivors Network" available at: https://www.counterextremism.org/resources/details/id/415/global-survivors-network accessed 4/10/2017

[20] Although outside the scope of this paper, where counter messaging campaigns are reliant on time-limited support from organisations, they can appear transient and short-term. This is a marked contrast with the larger and more persistent libraries of content that seem to be developed by both informal counter messengers, and extremists.

[21] Ferguson, *Countering Violent Extremism Through Media and Communication Strategies*.

[22] Glazzard, *Losing the Plot: Narrative, Counter-Narrative and Violent Extremism*.

[23] Tom Holt, Joshua Freilich, Steven Chermak, and Clark McCauley, "Political radicalization on the Internet: Extremist content, government control, and the power of victim and jihad videos," *Dynamics of Asymmetric Conflict*, *8*(2) (2015), pp. 107–120.

[24] Silverman, et al, *The Impact of Counter-Narratives: Insights From a Year-Long Cross-Platform Pilot Study of Counter-Narrative Curation, Targeting, Evaluation and Impact*.

[25] Stefan Malthaner & Peter Waldmann, "The Radical Milieu: Conceptualizing the Supportive Social Environment of Terrorist Groups," *Studies in Conflict & Terrorism*, 37(12) (2014), pp. 979-998.

[26] Anthony Richards, "From terrorism to 'radicalization' to 'extremism': counterterrorism imperative or loss of focus?" *International Affairs*, 91(2) (2015), pp. 371-380.

[27] Secretary of State for the Home Department, *Counter Extremism Strategy*, (October 2015) Available at: https://www.gov.uk/government/publications/counter-extremism-strategy accessed 05/12/17. See also: Martin Innes, Colin Roberts, and Trudy Lowe, "A Disruptive Influence? "Prevent-ing" Problems and Countering Violent Extremism Policy in Practice," *Law & Society Review* 51(2) (2017), pp. 252-281.

[28] Joint Committee on Human Rights, *Counter Extremism,* (July 2016). Available at: https://publications.parliament.uk/pa/jt201617/jtselect/jtrights/105/10502.htm accessed 05/12/17.

[29] Ferguson, *Countering Violent Extremism Through Media and Communication Strategies*.

[30] Steven A. Seidman, *Posters, Propaganda, and Persuasion in Election Campaigns Around the World and Through History*, (Oxford: Peter Lang, 2008).

[31] Edward Herman, and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media*, (New York: Vintage, 1995).

[32] Edward L. Bernays, *Propaganda*, (New York: Ig, 2005).

[33] John Corner, "Mediated Politics, Promotional Culture and the Idea of 'Propaganda'," *Media, Culture & Society* 29(4) (2007), pp. 669-677.

[34] BBC "Muslim Council says Prevent anti-terror scheme has 'failed'" (2014) available at: http://www.bbc.co.uk/news/uk-28934992 accessed 21/9/16.

[35] Ian Cobain, Alice Ross, Rob Evans, and Mona Mahmood, "Help for Syria: the 'aid campaign' secretly run by the UK government," (2016) available at: https://www.theguardian.com/world/2016/may/03/help-for-syria-aid-campaign-secretly-run-by-uk-government. Accessed 26 May 2017.

[36] Ben Hayes, and Asim Qureshi, *We Are Completely Independent: The Home Office, Breakthrough Media and the PREVENT Counter Narrative Industry*, (London: CAGE, 2016)

[37] Tim Aistrope, "The Muslim Paranoia Narrative in Counter-Radicalization Policy", *Critical Studies on Terrorism*, 9(2) (2016). Pp. 182-204.

[38] Michael Tierney, "Combating homegrown extremism: assessing common critiques and new approaches for CVE in North America," *Journal of Policing, Intelligence and Counter Terrorism*, 12(1) (2017), Pp.66-73.

[39] Briggs & Feve, *Review of Programs to Counter Narratives of Violent Extremism;* Ferguson, *Countering Violent Extremism Through Media and Communication Strategies*; Alex Schmid. *Al- Qaeda's "Single Narrative" and Attempts to Develop Counter- Narratives: The State of Knowledge*, (The Hague: International Centre for Counter Terrorism, 2014).

[40] Graeme Herd, and Anne Aldis. "Synthesizing Worldwide Experiences in Countering Ideological Support for Terrorism (CIST)" in Anne Aldis, and Graeme Herd (eds) *The Ideological War on Terror: Worldwide Strategies for Counter Terrorism*, (London: Routledge, 2016) p247.

[41] Schmid. *Al- Qaeda's "Single Narrative" and Attempts to Develop Counter-Narratives: The State of Knowledge*.

[42] Tierney, "Combating homegrown extremism: assessing common critiques and new approaches for CVE in North America".

[43] Coolnessofhind (2015) "Abdullah-X or Abdul-Neocon?" (2015) available at: https://coolnessofhind.wordpress.com/2015/05/06/abdullah-x-or-abdul-neocon/ accessed 4/10/ 2017.

[44] Herd & Aldis, "Synthesizing Worldwide Experiences in Countering Ideological Support for Terrorism (CIST)".

[45] Ashour,"Online De-Radicalization? Countering Violent Extremist Narratives: Message, Messenger and Media Strategy".

[46] Herd & Aldis, "Synthesizing Worldwide Experiences in Countering Ideological Support for Terrorism (CIST)"; Sarah Marsden. "Conceptualising "success" with those convicted of terrorism offences: Aims, methods, and barriers to reintegration," *Behavioral Sciences of Terrorism and Political Aggression*, *7*(2) (2015), pp.143–165.

[47] Jacqui True and Sri Eddyono. 'Preventing Violent Extremism: Gender Perspectives and Women's Roles'. Monash University, n.d.

[48] The original interview can be heard here: https://www.youtube.com/watch?v=kjuNuqIev8M and the auto-tune version can be heard here: https://www.youtube.com/watch?v=AIPD8qHhtVU

[49] J. Walsh "Britain Furst: the halal Ray-ban-wearing far right Facebook mockers." (2014) available at: https://www.theguardian.com/media/2014/jun/20/britain-furst-the-halal-ray-ban-wearing-far-right-facebook-mockers. Accessed 26 May 2017.

[50] Tanya Silverman, Christopher J. Stewart, Zahed Amanullah, Jonathan Birdwell, *The Impact of Counter-Narratives: Insights From a Year-Long Cross-Platform Pilot Study of Counter-Narrative Curation, Targeting, Evaluation and Impact*, (London: Institute for Strategic Dialogue, 2016).

[51] Jacob Davey, and Julia Ebner, *The Fringe Insurgency: Connectivity, Convergence and Mainstreaming of the Extreme Right*. London: Institute for Strategic Dialogue. (2017), p6.

[52] The Redirect Method "The Redirect Method: A Blueprint for Bypassing Extremism," available at: https://redirectmethod.org/ accessed: 19 October 2017.

[53] Laura Huey, Johnny Nhan, Ryan Broll 'Uppity Civilians' and 'Cyber-Vigilantes': The Role of the General Public in Policing Cyber-Crime" Criminology & Criminal Justice 13(1) (2012) pp. 81-97.

[54] Johnny Nhan, Laura Huey and Ryan Broll "Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings" British Journal of Criminology 57(2) (2015) pp. 341-361.

[55] Ibid. p.342

[56] Dupont, Benoît. 'Security in the Age of Networks'. *Policing and Society* 14, no. 1 (March 2004): 76–91.

[57] Huey et al, "'Uppity Civilians' and 'Cyber-Vigilantes': The Role of the General Public in Policing Cyber-Crime".

[58] Nhan et al, "Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings".

[59] Aistrope, "The Muslim Paranoia Narrative in Counter-Radicalization Policy".

[60] Adrian Cherney, "Designing and Implementing Programmes to Tackle Radicalization and Violent Extremism: Lessons from Criminology", *Dynamics of Asymmetric Conflict*, 9(1-3) (2016), pp. 82-94.

[61] C. Nadine Wathen, and Jacquelyn Burkell, "Believe It or Not: Factors Influencing Credibility on the Web," *Journal of the American Society for Information Science and Technology*, 53(2) (2002), pp. 134-144.

[62] Wathen & Burkell, "Believe It or Not: Factors Influencing Credibility on the Web".

[63] Jacques Ellul, *Propaganda: The Formation of Men's Attitudes,* (Vintage: New York, 1973).

[64] Aistrope, "The Muslim Paranoia Narrative in Counter-Radicalization Policy".

[65] Schmid. *Al- Qaeda's "Single Narrative" and Attempts to Develop Counter-Narratives: The State of Knowledge*.

[66] Herd & Aldis, "Synthesizing Worldwide Experiences in Countering Ideological Support for Terrorism (CIST)".

[67] Tierney, "Combating homegrown extremism: assessing common critiques and new approaches for CVE in North America".

[68] Richard Dawkins, *The Selfish Gene*, (Oxford: Oxford University Press, 1989); Limor Shifman, "Memes in a Digital World: Reconciling with a Conceptual Troublemaker," *Journal of Computer-Mediated Communication*, *18*(3) (2013), pp. 362–377.

[69] Shifman, "Memes in a Digital World: Reconciling with a Conceptual Troublemaker".

[70] Benjamin Lee and Vincent Campbell, Looking Out or Turning In?: Organizational Ramifications of Online Political Posters on Facebook" *The International Journal of Press/Politics*, 21(3) (2016), pp. 313-337.

[71] Laura Huey "This is Not Your Mother's Terrorism: Social Media, Online Radicalization and the Practice of Political Jamming," *Journal of Terrorism Research*, 6(2) (2015), pp. 1-16.

[72] For the full video, see: See: https://www.youtube.com/watch?v=o3VDDbh5dXw

[73] Greg Miller and Scott Higham, "In a Propaganda War Against ISIS, the US Tried to Play by the Enemies Rules," The Washington Post May 8 2015 available at: https://www.washingtonpost.com/world/national-security/in-a-propaganda-war-us-tried-to-play-by-the-enemys-rules/2015/05/08/6eb6b732-e52f-11e4-81ea-0649268f729e_story.html accessed 14 June 2017.

[74] See: https://www.youtube.com/watch?v=tMPfJJNeZUI

[75] Roger Eatwell, "Community Cohesion and Cumulative Extremism in Contemporary Britain," *The Political Quarterly*, 77(2) (2006), pp. 204-216.

[76] Joel Busher, and Graham Macklin. "Interpreting "Cumulative Extremism": Six Proposals for Enhancing Conceptual Clarity," *Terrorism and Political Violence,* 27(5), pp. 884-905.

[77] Jamie Bartlett and Alex Krasodomski-Jones, *Counter-speech Examining Content that Challenges Extremism Online*, (London: Demos, 2015).

[78] Benjamin Lee, "A Day in the "Swamp": Understanding Discourse in the Online Counter-Jihad Nebula," *Democracy & Security*, 11(3), pp 248-274.

[79] Sean Everton "Social Networks and Religious Violence," *Review of Religious Research*, 58(2) (2016) pp. 191-217

[80] The sketch can be viewed here: https://www.youtube.com/watch?v=_L2fazw5Y9k accessed 14 June 2017.

[81] Interview with research participant (2017).

[82] Zac Cochrane "Islamophobia Awareness Training in Schools" (2016) available at: http://uaf.org.uk/2016/07/islamophobia-awareness-training-in-schools/ accessed 19 June 2017.