



Countering Violent Extremism Online: The Experiences of Informal Counter Messaging Actors

Benjamin Lee 

The online space is a haven for extremists of all kinds. Although efforts to remove violent and extremist content are increasing, there is a widely accepted need to also contest extremist messages with counter messages designed to undermine and disrupt extremist narratives. While the majority of academic focus has been on large and well-funded efforts linked to governments, this article considers the experiences of informal actors who are active in contesting extremist messaging but who lack the support of large institutions. Informal actors come without some of the baggage that accompanies formal counter message campaigns, which have been attacked as lacking in credibility and constituting “just more government propaganda.” This has been noted by some of the wider countering violent extremism industry and the appetite for incorporating “real-world” content in their campaigns seems to be rising. This article fills a gap in our knowledge of the experiences of informal counter messaging actors. Through a series of in-depth qualitative interviews it demonstrates that, despite the potentially serious risks of incorporating greater levels of informal content, there is an appetite among informal actors to engage with formal campaigns where they can be selective over who they work with and maintain a degree of control.

KEY WORDS: counter messaging, counter narrative, terrorism, extremism, countering violent extremism, security

网络空间是所有极端主义者的避难所。尽管为消除暴力和极端内容的相关努力正在增加，但仍广泛需要使用那些用于破坏极端叙事的“反击信息”来对抗极端主义信息。尽管大多数学术一直聚焦于与政府有关的、资金充足的大型反极端主义信息工作，但本文研究了非正式行为者的经历。这些行为者积极反击极端主义信息，但却缺乏大型机构的支持。非正式行为者并没有那些伴随正式反击信息运动的包袱，比如缺乏可靠性和构建“更多的政府宣传”。“反击暴力极端主义”(CVE)的相关产业也注意到了这一点，因此将“现实世界”内容融入其宣传运动的势头似乎有所增加。本文填补了针对非正式信息反击行为者经历的研究空白。通过进行一系列深度定性访谈，本文证明，尽管融入更高层次的非正式信息会产生潜在严重风险，但非正式行为者有强烈愿望参与正式活动，在正式活动中，他们能选择与谁合作，并享有一定程度的控制权。

关键词： 信息反击，反击叙事，恐怖主义，极端主义，反击暴力极端主义，CVE，安全

El espacio en línea es un refugio para extremistas de todo tipo. Aunque los esfuerzos para eliminar el contenido violento y extremista están aumentando, existe una necesidad ampliamente aceptada de

impugnar los mensajes extremistas con mensajes en contra diseñados para socavar e interrumpir las narraciones extremistas. Si bien la mayor parte del enfoque académico se ha centrado en esfuerzos grandes y bien financiados vinculados a los gobiernos, este artículo considera las experiencias de los actores informales que participan activamente en la lucha contra los mensajes extremistas pero que carecen del apoyo de grandes instituciones. Los actores informales vienen sin parte de la carga que acompaña a las campañas formales de contra mensajes, que han sido atacadas por carecer de credibilidad y constituyen “solo más propaganda gubernamental”. Esto ha sido señalado por parte de la industria CVE y el apetito por incorporar “acciones reales” parece estar aumentando. Este artículo llena un vacío en nuestro conocimiento de las experiencias de los actores informales de contra mensajes. A través de una serie de entrevistas cualitativas en profundidad, demuestra que, a pesar de los riesgos potencialmente graves de incorporar mayores niveles de contenido informal, existe un interés entre los actores informales para participar en campañas formales en las que pueden ser selectivos con respecto a quiénes trabajan y mantienen un grado de control.

PALABRAS CLAVE: contra mensajes, contra narrativa, terrorismo, extremismo, contrarrestar el extremismo violento, CVE, seguridad

Introduction

While academic and policy uncertainty around terrorism and extremism has raged, there has been recognition that responses to terrorist violence cannot be limited to policing and security alone. This has given rise to a wide range of charities, governments, activists, and think-tanks aiming to get involved in the field that has become known as countering violent extremism, or more usually: CVE. Within this broad-ranging policy area, counter messages are attempts to negate the ideological draw of terrorism and extremism. None of these activities are simple, and there has been a great deal of criticism of CVE broadly, and counter messaging specifically. Chief among the concerns has been the ability of counter messaging providers to deliver convincing and authoritative counter messages. Put simply, why would audiences listen to a word the counter messaging “industry” has to say?

Counter messaging providers are aware of the difficulties they face, and so the idea of importing non-professional “real world” counter messaging content into campaigns has taken hold with some. Produced by amateurs with no (or at least different) axes to grind, informally produced content has the potential to be more authentic and, it is often assumed, persuasive. This is largely down to the perceived greater credibility and authenticity of informal actors compared to those embedded in larger organisations with ties to the state (Lee, 2019). But little is known about informal counter messaging activists and their motivations, their experiences, and crucially their attitudes to professional counter-messaging provision. Using a limited sample of informal actors¹ drawn mainly from the United Kingdom, this article seeks to gain some insight into the motivations and methods of those who seek to produce counter messages without the support of larger organizations. It also aims to evaluate as far as possible the potential for collaboration between professional and informal actors.

What follows is a brief discussion of the concept of counter messaging, and a summary of the main criticisms of counter messaging approaches to counter terrorism, focusing on the question of message credibility. This debate is then expanded to encompass recent statements from counter messaging providers that suggest they intend to make greater use of informally produced counter messaging content in future campaigns. The virtues of this approach are discussed followed by an outline of the main research questions considered in this analysis.

Counter Messaging and Counter Messengers

Counter messages aim to disrupt the communicative activities of terrorist and extremist political and religious groups, either by undermining extremist messengers or messages, or by convincing audiences with alternative ideological messages. There has been no universal agreement on the balance between counter and alternative messages, the focus on specific messages or broader narratives, or the exact audiences targeted (Ferguson, 2016; Glazzard, 2017; Lee, 2019).

A key dimension of variation has been between strategies that seek to tear down extreme ideological positions and those that seek to instill their own. Briggs and Feve (2013), in a report authored by the think-tank Institute for Strategic Dialogue (ISD), have differentiated between alternative narratives that seek to tell positive stories, and counter-narratives that attempt to challenge and discredit existing extremist narratives. Writing on online radicalization in 2013, Peter Neumann argued that measures should be taken to “reduce the demand” for violent extremist messaging, including “by discrediting, countering, and confronting extremist narratives or by educating young people to question the messages they see online” (Neumann, 2013, p. 433). Into this space has also seeped the use of humor and ridicule to undermine the credibility of extremist messages, a long-running strategy in evidence since the U.S. War of Independence, and used notably in the Second World War (Goodall et al., 2012, p. 72). In contrast, other accounts have linked counter narratives more explicitly to a “battle of ideas” and propaganda (Payne, 2009), suggesting that the end of jihadist ideology is likely to bring about the end of jihadist terrorism (Sedgwick, 2012, p. 359).

There is further variation still between analyses that concentrate on messages, and those that concentrate on the concept of narrative. For example, Braddock and Horgan (2016, p. 383–384) stress the importance of narrative over simple messages; they argue that narratives are more persuasive, as they superficially set out to inform and entertain rather than persuade, thereby negotiating the traditional barriers to persuasion. Subsequent work by Braddock and Morrison (2018) has further developed their analysis around the cues and heuristics of online communication and how they can affect perceptions of trustworthiness. Glazzard (2017) argues against devoting resources to counter narratives, in part because they are based on what he considers weak theoretical foundations. The term narrative, Glazzard (2017, p. 8) argues, is used interchangeably with ideology, while insufficient attention has been paid to the term’s literary origins. Whether the distinction between narrative, ideology, and message is fully realized within CVE

policy and counter messaging is highly debatable. For the sake of inclusivity, this article focuses solely on counter messaging, defined as aiming to contradict specific claims, rather than the more specific concept of narrative.

A final point of variation has been on the basis of audiences. In much of the counter messaging literature it is unclear who is being targeted by attempts to discredit extremist messages, or to lure away its adherents. CVE policy more broadly can be defined as falling into one of three categories depending on the target (CREST, 2018). Primary CVE (sometimes referred to as Preventing Violent Extremism or PVE) aims to target broad audiences before radicalization takes place. Secondary CVE focuses in on those considered to be at risk—in medical language, showing symptoms of extremism. Tertiary CVE engages with those already involved in extremism or violent extremism (depending on the policy), for example, working through exit-style programs (Harris-Hogan et al., 2016).

Counter messaging can also be viewed in these terms depending on the audience targeted. Many attempts at counter messaging appear to be generalized, producing content with seemingly little attempt to direct it to specific audiences. These include some of the well-known officially-backed counter messaging campaigns including *Average Mohammed* and *Abdullah X*, both of which appear as online portals to counter messaging content. More direct interventions are closer to secondary or tertiary CVE, seeking to engage in conversations with those already involved or on the periphery. Braddock and Horgan (2016, p. 387), for example, advocate targeting “those at risk of radicalization.” Programs that take this approach seek to identify and target social media users that demonstrate some sort of risk factors, and to contact them directly (Frenett & Dow, 2015).

Criticisms of counter messaging have included questioning its strategic effectiveness, as well as wider concerns about the credibility of counter messaging strategies. A 2016 review of media and communication strategies within CVE policy noted the largely unknown role of narrative and propaganda in explaining involvement in violent extremism, and therefore the unknown consequences of counter narrative (Ferguson, 2016, p. 9–15). Other accounts have also noted the weak evidence base for the role of counter narratives in dissuading participation in violent extremism and questioned the basis for government and civil society enthusiasm for the approach (Glazzard, 2017, p. 5). Some have also suggested that, at its worst, counter messaging can even be counter-productive (Bartlett & Krasodonski-Jones, 2015, p. 5). While the empirical basis for counter messaging is undeniably shaky, much of this critique rests on varying interpretations of what counter messages are designed to achieve. While the dissuasion of audiences already on the periphery or involved in violent extremist milieus is unlikely, little focus has been given by critics to audiences that are not yet connected with violent extremism. Even if it fails to tempt adherents away from violent extremist ideologies, counter messaging is likely to maintain societal rejection of violent extremist narratives.

Beyond these concerns is the question of credibility. This critique is partially normative, with much of the criticism centered on the idea that CVE policy and counter messaging are overreactions, and that they represent propagandistic state

attempts to stifle debate and unpalatable political beliefs (Aistrophe, 2016, p. 17). Criticism has been heightened where states have acted covertly. In the United Kingdom, a Home Office backed campaign, *Help for Syria*, provoked a backlash when it distributed leaflets without acknowledging its government support (Cobain et al., 2016). As well as the idea that governments shouldn't be involved in counter messaging, questions of credibility also raise the idea that governments and the organizations they support are hopelessly compromised in the eyes of target audiences, and therefore incapable of delivering effective counter messages (Braddock & Horgan, 2016; p. 386; Schmid, 2014, p. 14). In part, this stems from widely perceived gaps between government policy and narratives (Aistrophe, 2016). Briggs and Feve (2013, p. 16) have argued that a government "credibility gap" means that governments should limit themselves to enabling other civil society actors to engage in counter messaging work. Much of the literature, and most of the criticism, has focused on government and government-backed counter messaging campaigns. However, even where campaigns are not delivered by government agencies, they are often carried out by organizations that receive support from governments:

While governments have attempted to include civil society and the private sector in efforts, there is still a fear that governments are the ultimate drivers of these programs. That fear can then lead to mistrust amongst the intended recipients of CVE campaigns, not least of which because at the same time, governments are also looking to collect useful information to lay charges and make arrests. (Tierney, 2017, p. 70)

Informal Counter Messaging—Why Does Everything Have to be a Campaign?

To summarize so far, attempts to counter extremist messages and narratives initiated by states have been critiqued on the grounds of lack of effectiveness, in part stemming from the unsuitability of state and supported actors to drive counter messaging campaigns. The argument for unsuitability originates from a normative belief that states should not engage in propagandistic ideological contests, and that the behavior of many states has hopelessly compromised them in the eyes of many target audiences. This lack of credibility can also be assumed to apply to state-backed civil society organizations.

However, these criticisms, and accounts of counter messaging more generally, do not address the spontaneous and informal production of counter messages inherent in societies. Extreme messages are, by definition, at odds with mainstream views. While governments have been seeking to produce and stimulate campaigns that delegitimize extremist messengers, the general public are already engaging in these behaviors informally without any impetus from either states or civil society organizations. Much of this counter messaging is likely to occur in opaque settings, potentially at the interpersonal level, for example in conversations with friends, or around the kitchen table with family. There is emerging research highlighting the

contributions of friends and family in preventing violent extremism (True & Eddyono, 2018).

More easily accessible are the many examples of individuals using digital media to publicly critique and in many cases ridicule extremist messages, often with greater apparent success than formally organized campaigns. Prominent examples include the widespread ridicule of the English Defence League (EDL), an extremist counter-jihadist street group, following the production of a YouTube video featuring music with lyrics taken from a drunken interview with an EDL supporter. The resulting *Muslimic Ray Guns* meme, mocking the incoherent interviewee and playing up to the stereotype of EDL supporters as uneducated, has inspired multiple other parodies and is available to be purchased as a tee-shirt.² A further example was the commercially produced film *4 Lions*. Directed and co-written by satirist Christopher Morris, the film depicted an inept band of Yorkshire-based would-be jihadists in a plot that draws heavily on the attacks on the London transport network in 2005. Less prominent examples of informal counter messaging include the emergence of parody accounts on social media such as the sub-reddit *Behold the Master Race*,³ and the Twitter account @mock_the_right.⁴ Each of these examples easily fits into the realm of primary CVE and in effect seeks to maintain the social prohibition on extreme ideas and behaviors.

The extent to which the informal and formal production of counter messages can align remains an open question. There has been appetite from formal counter messaging organizations to tap into informally produced content available online. In 2016, the think-tank ISD recommended that “natural world” content could be used in campaigns (Silverman et al., 2016). The “Re-direct method,” an online blueprint for “bypassing extremism” that lists Google’s incubator Jigsaw as a supporter, 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 [sic] rejected outright by our target audience.*⁵

Another innovation has come in the form of The Disruptors. Encouraging visitors to “Hack the narrative,” The Disruptors is a collaboration between three organizations (RLLeaders, America Abroad Media, and Affinis Labs). Most recently The Disruptors ran a competition in which the public were encouraged to submit “original memes to counter violent hate” with the promise of cash prizes for those memes that are shared most successfully.⁶

There is also a strong theoretical argument that predicts closer working between citizens and security providers. Benoît Dupont has attempted to conceptualize security as emerging from complex networks rather than through the division of state and market forces. Using capital as a metaphor, he argued that each actor within a security network was able to mobilize their own resources to

solve a security problem (Dupont, 2004). This idea was taken further in the context of digital security, with “uppity civilians” able to insert themselves into security networks by virtue of their skills and resources (Huey et al., 2013). Perhaps the most striking example of this was the crowd-sourced investigation into the Boston Marathon Bombing. Despite the failure of the crowd-sourced investigation to identify the perpetrator, in the words of one researcher: the “genie is out of the bottle” (Nhan et al., 2017, p. 2). Increasingly, it seems, private citizens with capital, measured by skill, insight, or some other resource lacking in official agencies, are looking to contribute to solving security problems. As the initiatives by ISD and the Disruptors indicate, it makes sense in the digital environment for larger, more formal organizations that are seeking to generate counter messaging content, to go to informal actors. This both allows them to circumnavigate the credibility issues attached to “official” content, and to borrow some of the insight and cultural capital available to informal actors enmeshed in digital cultures.

A previous theoretical analysis (Lee, 2019) has identified the potential risks stemming from cooperation between formal and informal counter messaging actors from the perspective of CVE; among them:

- Contrasting motivations for content production between formal and informal actors.
- Unknown attitudes toward collaboration with larger organizations among informal actors.
- Diverse approaches to content creation, including the use of humor, offence, and the ridicule of individuals and specific organizations that would be unpalatable for larger organizations.
- Informal messengers may be exposed to greater personal risk (online and offline) than those working for formal organizations. Closer working may see larger organizations effectively exploiting the willingness of informal actors to take risks with their online and offline safety.

Overall, it was suggested that, despite the arguments for greater collaboration between formal and informal actors, little was known about informal actors and their motivations. This research is a first step in getting to grips with informal counter messaging, by examining the experiences of informal counter messaging actors who are active online. Understanding their motivations, experiences, and attitudes to *formal* counter messaging providers is a crucial component in assessing the viability of closer working between formal and informal counter messaging content providers.

Data and methods

Where this article seeks to contribute is in making an initial attempt to understand the experiences and motivations of informal counter messaging actors. This is a first step to understanding what kind of messages are being disseminated, to what end, what the reactions have been, and what the potential for future

Table 1. Participants and Descriptions

Participant	Description
1	A male producing videos disseminated on YouTube, Facebook, and Twitter to a small audience. Videos often include elaborate parodies of Islamist extremist figures as well as members of the far-right and other, non-extremist figures.
2	A male who has previously run a large Facebook account parodying a U.K. far-right group.
3	A male with a film school background producing audio and video content disseminated over social media and the web to a small audience. Content is broad-ranging but a small number of videos have been highly critical of a U.K.-based far-right group.
4	A male with a background in wider countering violent extremism (CVE) work running a mid-sized Facebook group designed to speak out against Daesh/Islamic State.
5	A male who runs a Twitter account that aims to disseminate statements made by members of a U.K.-based far-right group in an attempt to discredit them.
6	A male living in Israel who produces videos focused on public opinion.
7	A female running a Facebook page opposing far-right extremism with a focus in her local area.
8	A female jointly running a large anti-extremism group with others with a Twitter account as well as other forms of social media activism, opposing far-right extremism with a national focus.
9	A female running a modest anti-extremism group mainly through a Facebook page as well as other forms of social media activism, opposing far-right extremism with a national focus.

collaborations between informal and formal sectors might be. This article addresses four related research questions:

- 1) What motivates informal counter-messaging content actors to produce their content?
- 2) How are informal counter messaging actors attempting to intervene and what strategies do they see themselves as pursuing?
- 3) What have been their experiences of producing content, including any potential hostility and risks?
- 4) How do they feel about potential closer working with formal counter messaging campaigns?

This article draws on nine interviews with informal counter messaging activists (Table 1). Some level of abstraction has been necessary to preserve confidentiality.

All interviews were carried out between December 2016 and December 2017. Sampling was purposeful and potential participants were approached on the basis of counter messaging content that the researcher was able to identify. As the subject of study was counter messaging defined broadly, participants were not selected on the basis of their focus on countering specific forms of extremism. The sample is split between those opposing the far-right and those opposing Islamist extremism. The researcher did not focus specifically on content that he believed to be suitable for use in larger scale counter messaging campaigns. Several of those approached declined to take part in the research project. Reasons cited included skepticism over the research project as well as the length of time that had passed following the production of content. In other cases, there was no response to contact attempts. It

was unclear if non-response represented active rejection of participation or if messages were not received.

Interviews were conducted in-person using a semi-structured format with the exceptions of participants 6 and 9, where interviews were conducted over Skype and email, respectively. To protect participant confidentiality interviews were transcribed solely by the researcher and the completed transcripts were de-identified. The need to preserve participant confidentiality, combined with the public availability of the content produced by participants means that explicit references to content produced by participants were removed to prevent identification. The total amount of transcribed material amounted to slightly under 75,000 words. Analysis was thematic and in large part deductive. The framework for analysis was based on categories derived from the research questions which, in turn, were based on substantial theoretical work on the relevance of informal counter messaging to larger, more organized efforts (see discussion above, and Lee, 2019). However, semi-structured interviews also allowed for interesting findings to emerge deductively, and for the frame work to be partially adapted in response to the data.

There were limitations to this method. The use of non-systematic sampling raises the issue of selection bias. This is to some extent an inevitable issue but caution should be exercised in any attempt to draw widely generalizable conclusions from this data. The sample is likely exceptional in many respects. First, the researcher's background is as a U.K.-resident, native English-speaker with a research interest in the extreme right. This makes English-language content addressing the extreme right easier to access and comprehend for the researcher. Although attempts were made to include counter messages addressed towards religious extremists, notably Islamist extremism, the bulk of the participants (six of nine) were engaged in producing content against the extreme right. One participant was engaged exclusively in targeting violent Islamism, and the final two participants were targeting both the extreme right and Islamist extremism to some degree. A further limitation is the focus on publicly available social media. A major contention of this research is that counter messaging occurs at multiple levels including through interpersonal communication. Social media represents an easily available but narrow slice of possible routes for counter messaging to reach relevant audiences.

The researcher was open about his research and the funding organizations supporting it (the Centre for Research and Evidence on Security Threats, CREST, is funded in part by the U.K. security and intelligence agencies). This may have contributed to non-response by some potential participants, in particular those who were critical of the U.K. government. As a result, the sample may underrepresent some of the more ideologically driven actors, although some of the interviewees were skeptical of U.K. policy.

Finally, the overall sample size is small. Increasing the sample size would be dependent on greater resources. Likewise, further expertise, for example, in religious extremism, or other languages, would also allow for greater diversity in

the sample. Currently, the data available cannot by any stretch of the imagination be considered the final word on this subject. Although the choice of methods is appropriate and the resulting data rich and insightful, sampling issues mean that this article represents only a narrow set of perspectives. Further research will be needed to develop the conclusions offered here.

The Experiences of Informal Counter Messaging Actors

Motivation

On the basis of the available material, very few of the participants described their behavior as being ideologically driven or linked to explicit political activism. For the most part, the participants were not seeking to promote their own cause, only to question the extreme narratives they were targeting. In some instances, behaviors were consciously decoupled from any political agenda:

Just because. It was fucking funny. I had the time and resources to do it and I could, so I did. (Participant 2)

Some participants rooted their activity more explicitly in the realm of politics, with several participants opposing the far-right concerned about wider societal trends stemming from political and economic malaise:

It's always after a downturn, after recessions and stuff, the people rise up and people get divided. (Participant 3)

I've studied a lot of the rise of national socialism in Germany, I think you can draw a lot of parallels. From what I can see it gains popularity through times of economic hardship. (Participant 5)

Motivations for activity were frequently more readily described in the form of personal anecdotes. As well as the historical context, participant 5 described their experience of interacting with a far-right group online:

I commented. I used my own personal account and said this was wrong, this is why. And they banned me. My Facebook account, I can't comment on [far-right page]. That really pissed me off. (Participant 5)

Participant 6 also attributes the decision to begin producing counter messages to an experience online and a desire to demonstrate the diversity of opinion on a topic, rather than to advance a specific position:

At a certain point I was bored, and somebody was making a lot of claims about Israel and Israelis, and it kind of annoyed me. You can claim a lot of things, but at least have some evidence. (Participant 6)

Participant 4 listed a mix of concerns, from the general, to a specific campaign issue. Generally, participant 4 was seeking to amplify Muslim condemnation of violent Islamist extremism:

Why are X number of Muslims actually speaking out against terrorism, and why is it not being picked up by the wider masses? (Participant 4)

but also with a narrow goal within that broader aim to lobby against the use of the term Islamic State:

...one of our earlier aims was to ask, or to lobby, politicians and the BBC to stop using the term Islamic State and start using the term Daesh. (Participant 4)

Participant 8 linked her activism to her online opposition to an individual far-right Twitter account she saw as building a following. From there she describes how these activities coalesced into a more concrete counter messaging organization:

There was no cohesive structure to it. It was literally people on Twitter getting together and organizing reporting sessions against these people, arguing with people online to try and get people to see their message, and to try and put people off joining the organization. (Participant 8)

Participant 9 describes her experiences of watching the success of far-right click bait with notionally left-wing friends, with a sense of personal affront originating from the perceived deceptive practices of the far-right:

So...my left-wing friends, I felt, were being tricked into sharing (promoting) [far-right group] and being tricked into donating funds to the [organization], as were millions of others. I was absolutely outraged. Sure, if you want to donate to them and promote them, fine, but do it because that's what you're choosing to do, not because you've been hoodwinked into it. (Participant 9)

Participant 1 stands out as different from the other cases. Participant 1 saw his content as rooted in disaffection with the wider state of affairs within the Muslim community. Opposition to extremism was only a single component of a wider reformist agenda for him. He describes a frustrating upbringing in a religious school as a catalyst for the production of content:

It almost felt like I was the only guy in the room who finds this [religion] funny. (Participant 1)

Participant 1 chose to release content parodying what he saw as Islamist extremist preachers out of a similar disbelief that others took it seriously:

That's why I released [video parodying an Islamist preacher] that's got over 1,000,000 views. That's because a lot of people actually think the guy's a bit of a clown, he's a bit of a fraud. But if you were to ask most people on the streets, they probably wouldn't say, but secretly they probably think he's a bit of a buffoon. (Participant 1)

Opposing extremist messages is, to some extent, inherently political. However, there was little indication in the sample that counter messaging content was a component in a wider ideological program. Instead, informal actors were often motivated out of a sense of personal grievance, linking their activism to specific incidents rather than wider ideological goals. Some linked their opposition to wider trends, others to nothing at all. The overall non-politicization of participants is further supported by their rejecting in many cases more traditional forms of opposition such as physical protest, as discussed below.

Strategies

As already noted, accounts of counter messaging vary in their conceptualization of counter messages, and this variation includes target audiences. Participants suggested a range of audiences for their work, including those supportive of extremist viewpoints and those already opposed. Many of those interviewed readily accepted that the messages they were producing would be of little value, or in some cases actually counter-productive, for those already accepting of extremist narratives:

I knew I was never going to make a difference to the people of the [far-right group], I think they're that far down the road. People like me, I don't think I can make a massive difference to them. (Participant 7)

Do I think I worked as a recruiting sergeant? Probably yeah. I think the people, the more articulate right-wingers I upset because they thought I was undermining the security of the nation. I think I probably made them more fervent in their beliefs. I probably reaffirmed people from a very liberal perspective: 'anti-far-right, I'm really cool.' (Participant 2)

In these cases, counter messaging content was often seen as important as a tool for awareness raising in the wider populace rather than as a tool for targeting and dissuading extremists. Participants often described their concerns that others should be made aware of extremist activism and if possible be urged to speak out:

I'm here to expose something I believe is awful and to try and do it in a humorous way that can expose it for what it really is. (Participant 5)

"It was really to get that sort of message that this is what was happening, and ordinary people needed to not pretend that it wasn't. Or even to not know it [...]"

was, and ideally not just ignore it but sort of say we don't agree with this, this is not OK." (Participant 7)

As the participants were active on social media, an emergent theme was the use of social media to reach beyond immediate audiences. While some participants acknowledged that their audiences were those who already agreed with them, they believed that viewers would share material more widely, thereby exposing their content to wider audiences:

But social media, the way social media works, is the people who will actually go on into our website and look at our articles are people who share our ideology. But the idea being then if they share it on platforms like Twitter, like Instagram, like Facebook where it will go to an audience where they may be on that marginal line where they don't have any defined ideology when it comes to anti-hate. (Participant 8)

A second perspective came from those who were more confrontational, with content designed and disseminated in such a way as to directly confront extremist groups. Participant 3 described creating a playlist on YouTube of far-right videos, and inserting his own anti far-right video:

I want everyone to see it. I want people to laugh at it. I want people to like share that stuff to see it. And they have watched it because there are six dislikes. And I put it in a playlist amongst their [far-right group] stuff. So you've got fifty videos of all their stuff, and then it's just in the middle. (Participant 3)

In some of these cases counter messaging was tied to other more confrontational tactics:

We started out very undercover. Doing very much the doxing work [publishing private identifiable information], the activism work. Then we've gone through this branding phase. And now we're becoming well known. (Participant 8)

In this context, counter messaging content was seen as a component in a wider project of inoculating the general population against extremist messages and making it difficult for extremists to advocate their ideas publicly.

It's more really about creating a climate of fear about expressing those opinions in the first place. We're savvy enough to know that there's so much hate out there that taking down one or two accounts isn't enough to make much of a difference. But what it does do is make other people think, 'Hang on, if I put that information out there that might happen to my account.' So it's fear and it's education through fear that we're looking at there. So we'll make examples of people. They know damn well that it's us that's done it while we're not openly saying it. The idea being that

people will think twice and that will stop the hate rather than just one individual person. (Participant 8)

Just as definitions and approaches to counter messaging have varied in academic literature, there have been a range of approaches on display amongst participants. Participants were, for the most part, of the view that counter messaging was unlikely to alter minds already made up. Their main audiences were those without extreme beliefs, with the intention of escalating the social costs of professing extreme beliefs. In some cases, counter messaging content was linked to more confrontational activities, including attempts to “out” extremist supporters by posting private information.

Risk

Participants described the risks that they undertook as part of their activities and there was a general acknowledgement that risks were associated with counter messaging.

I think for some there is a concern of safety. And that's something that has hit a chord with me. Particularly by other content creators, when it comes to counter narratives. So, I've had friends who have been threatened based on some of their counter narrative work. I haven't. I've been quite fortunate in that respect. (Participant 4)

Risk was generally seen as originating from the groups opposed in counter messages. This was largely confined to the online realm.

I think that's where I was a bit naïve at the beginning really, because I didn't realize there were people who were that violent and that aggressive, and that totally fixed, that they would speak to people openly on social media and threaten them with all kinds. You'd be there trying to have a civil, relatively civil conversation, or a debate, and you're being called a... Well it's not really a threat to be called a traitor, but start off with that, and once you challenge that and say what are you talking about? Well you're ashamed of your race. What? It'd be like, do you even know what you're saying? And I got threatened with being raped. That they'd find me. That they'd come to the house. I wouldn't say I got massively shook up by it, but I did say 'you're being a bit of an idiot.' (Participant 7)

Responses to threats varied. In some cases, participants took steps to publicize the threats that they received as a way of mocking those issuing them by highlighting their lack of credibility.

I don't get threats very often, and the threats that I do get I post them on Facebook because I want to laugh at them. Because they are usually pretty ridiculous, it's someone in Pakistan or Egypt who says they're going to come slit my throat. And

I'm like, how are you going to find me. You don't know where I live.
(Participant 6)

Surprisingly however, some participants reported threats that they had received from fellow campaigners. In particular, participants engaged in counter messaging against far-right organizations reported confrontations with more aggressive left-wing activists described as antifa. One interviewee describes encountering what she describes as an anti-fascist in an online support group for page administrators:

She's a nut job, she's an absolute... She is as bad as any far-right on that side, ever will be. I quickly fell out with her, not on purpose, but because I wasn't extreme and she didn't like it. And I'm very anti, I'm anti-demo, that type of thing, full stop. She was another one, I then got called all the things under the sun from. Talk about being stuck in the bloody middle. (Participant 7)

Another interviewee reported a similar experience:

I'm not friends with any of that extreme antifa group anymore. They're an unpredictable and dangerous group. (Participant 9)

None of the participants engaged in counter messaging against the far-right recognized themselves as being members of the anti-fascist movement, although it was difficult to tell if this was linked specifically to militant anti-fascism expressed sometimes through violent street protest, or the wider anti-fascist movement. One participant also reported risks arising from disputes within the group. In one incident, a disgruntled former group member was said to have spread rumors about the participant after leaving a group over a dispute:

He left after a bruised ego over a post which the team decided not to post and he wanted to post it. The team had said no, so that was the agreed decision. He has caused a lot of trouble in antifa circles, spreading rumors about me...
(Participant 9)

Not all risks were limited to the participants. Participant 7 reported her experiences with an extreme-right frequent visitor to her page. In this instance the participant ultimately banned the visitor because they were known to be a minor and were receiving abuse through the page as a result of their comments.

When he was 14, [I] blocked him, because he was a minor and I didn't want to get involved with... What he was saying was unacceptable really, were he an adult I'd probably, the whole free speech thing, I'd probably let it roll and roll. But he was starting to get attacked back and I was like I'm not comfortable with this, so I blocked him. But he came back, they just change their name and come back don't they. It was obvious it was him. (Participant 7)

In summary, most of the participants reported feeling under some kind of threat at some point. The nature of the threat varied. In some cases this was confined to online abuse and threats, which some (but not all) participants reported as having little overall impact on them. The choice to remain anonymous in many cases seems to stem from the recognition of how polarized the online environment was. Several participants also reported risks arising from unexpected sources. While the participants often chose to root their behaviors outside of explicit left/right political divisions, the presence of more militant left-wing activism against far-right groups in particular has exposed informal messengers to unanticipated risks. A further issue touched on in the interviews is the risk to extremists posting online as a result of counter messaging activism. In at least one instance a participant acknowledged that some users had taken advantage of her page to verbally attack an extremist supporter, and another openly talked of exposing extremist supporters publicly. This is consistent with other reports of escalations around counter messaging content, with counter messages and related comments sometimes coming to resemble hate speech (Bartlett & Krasodomski-Jones, 2015; Ernst et al., 2017).

This raises an important point for informal counter messaging actors. The absence of strong ideological motivations amongst the participants in this research suggests that the sample is drawn from non-militant actors. As Eatwell (2006) noted in his accounts of the 2001 Milltown riots in the United Kingdom, different forms of extreme politics and religion have been observed to feed each other. Militant left-wing activists are also engaged in counter messaging and other forms of direct action online. Equally, those opposed to Islamist extremism can also overreact to threats, rooting their activism in paranoid and conspiratorial Islamization narratives (Lee, 2016). The data gathered for this project suggests that many non-militant actors find themselves trapped between the form of the extremism they oppose, and the emergent alternate extremist position. There is a legitimate debate over the severity of harms arising from left-wing, right-wing, and Islamist extremisms, but for those engaging in counter messaging online the risks can emerge from both those they set out to oppose, and those who believe their opposition does not go far enough.

Collaboration

The major issue raised by this article is the attitudes of informal counter messaging actors towards collaboration with larger formalized organizations. This question follows on from the empirical and theoretical suggestions that formal organizations may seek to use content produced by informal actors. The surprising finding from this research was the extent to which the interviewees had already been approached by larger organizations. This has yielded mixed results. In some cases, relationships established with such organizations were viewed positively, in particular where an interviewee aspired to expand their presence. However, it was evident that care had been taken in selecting a group to collaborate with, and there needed to be some level of goal convergence for the interviewee to be comfortable.

I think the reason we like [larger organization], and I've got a lot of time for [larger organization], is that they are very very ideologically neutral organization. What I've found is a lot of pro-Muslim groups have an agenda of anti-Semitism, and a lot of Jewish groups vice-versa. [Participant's group] are completely about equality. (Participant 8)

The extent to which informal actors were picky about whom they partnered with was evident in other interviews too:

I have been approached by other organizations, wanting to use [group] to promote their counter narratives. And they've offered sweeteners at times, such as attendance at conferences in very nice locations, in five-star hotels, which I've refused. For me it has to be a legitimate part of me rather than somebody else then capitalizing on that. (Participant 4)

In one instance, overtures by larger organizations were interpreted as outright sinister:

It was quite young hip and trendy and they are very into getting young hip and trendy people into backing ideologies. I think. Literally at the time it was. And because there was an element of humor to it people don't necessarily think that there's some kind of underlying agenda behind humor, or anything like that. Especially that's supposed to be anti far-right, you know. There would have been a lot of pro-Islamic groups, fucking really into funding all that, that's why they fund [redacted]. They will take money from Palestinian interest. (Participant 2)

In other cases, attitudes to collaboration were more open. This was particularly the case for smaller actors or those seeking support in some form. When one actor was asked about the possibility of receiving training, for example:

That would be brilliant. Social media, or editing, how to promote. I'm not even talking about give me a thousand pounds... How to do basic things well. I think my channel suffers a lot from that. It could be more subscribers, more views. But because my stuff isn't maybe that professional level of editing or promotion. Or when I tweet something. I don't have that. I'm just learning on the go. That would be great. (Participant 1)

However, even in these instances actors indicated an ongoing concern about how content would be used, suggesting that a level of control would be important in any future collaboration.

For doing the right thing, I don't have a problem. Then again you do cross the line into 'Is it propaganda?' Government sponsored stuff is a bit whack. It all depends on what the content is and how it is going to be used and how you are going to come off in the end. (Participant 3)

In the terms of networked approaches to security (see above), there was evidence that, for some of the participants, collaborating with larger and more formal organizations could provide them with access to resources that would be of benefit. However, there was also a strong desire to retain some measure of control over content, indicating the risks to credibility that would come from being associated with non-impartial organizations. This seems particularly important in light of the largely politically neutral stance taken by the interviewees. While none of the interviewees explicitly acknowledged the symbolic or cultural capital they possessed, there was an implicit understanding that forming alliances with more formal organizations could potentially jeopardize it. More surprising, however, was the degree to which participants reported having already been approached by larger, more formalized organizations interested in collaboration. Some interpreted this interest as a sinister and cynical attempt to co-opt their content for political ends. Others were keen to develop alliances, but only with the right partners.

Conclusion—Battlefield Facebook

In an idealized theoretical world, there is a strong connection between the informal and formal messengers engaged in counter messaging. Asking informal actors to describe their experiences, however, suggests that any security networks built on collaboration between formal and informal actors are likely to be characterized by frictions.

One potential source of friction has been previously identified as motivation and ideology (Lee, 2019). Where strong political values are a component in counter messaging then its role may shift towards what some have conceptualized as alternative messaging, i.e. the advocacy of a different set of values. Worse, strongly ideological actors may attack opponents to demonstrate their own positions rather than to effect any positive social change; counter messaging becomes about proving a point rather than winning audiences over. Among the participants there was in most cases a sense of politics and ideological belief; however, this was rarely if ever seen as central to counter messaging. Most stressed the political inclusiveness of their campaigns or their overall lack of ideology. Based on this sample, this suggests that, in many cases, actors would be receptive to closer working with larger organizations.

Likewise, perspectives on counter messaging strategy were also encouraging from the perspective of formal–informal collaboration. Participants largely focused on reaching audiences that were already supportive of their views. Counter messages, in this instance, were seemingly designed to shore up the social prohibitions on extremist activism and to reinforce the illegitimacy of extreme messages. However, in some cases this was combined with forms of direct action that were more confrontational. In the case of Participant 8, counter messaging was combined with forms of social media-based direct action that would likely represent a severe reputational risk for any publicly supported organization.

More difficult for larger organizations to manage are the risks taken by informal counter messaging actors. Abuse, physical and legal threats, and

attempted public exposure of anonymous accounts are part and parcel of informal counter messaging. Interestingly, threats did not come exclusively from those extremist groups who were opposed, but also included more ideologically driven actors notionally on the same side. In one case, there was also a strong sense of responsibility towards those with extreme viewpoints, particularly minors, and the potential for them to receive abuse as a result of counter messaging activity. This supports the wider perspective of moderate counter messaging actors finding themselves trapped in an ideological middle ground, sandwiched between extremes. These risks would need to be considered in any closer cooperation with larger organizations. Failure to do so may result in large, well-resourced organizations outsourcing the risks of counter messaging onto well-meaning but largely unprotected actors. This is both a reputational risk and more importantly a moral hazard.

Despite these frictions, there is clearly potential for formal and informal organizations to work more closely. The best evidence for this is that many of the participants were either already actively working with larger organizations or had been previously approached. The fact that many of these approaches had been turned down indicates that informal actors are likely to be selective in who they work with. This may in part extend from an ideological commitment, but participants also talked about the need to maintain their own credibility—an important resource—as well as some degree of control over their content and how it is used. Despite this, there was a clear recognition of the resources possessed by larger organizations and how they could benefit informal actors, most notably in the form of training and promotion. In conclusion, in many cases there are good prospects for closer collaboration between formal and informal counter messaging actors. These need to be carefully considered, however, and only entered into with a clear understanding of the objectives of both parties.

A further point, unrelated to the original theoretical framework, concerns the role of social media platforms in facilitating both extremism and counter messaging directed against it. All of the participants were active on at least one of three social media platforms: Facebook, Twitter, and YouTube. To some extent this is to be expected as (i) social media represent the easiest and cheapest methods to distribute content online, (ii) the connectivity built into social media provides ready-made audiences, (iii) extremist groups are already active on social media, and (iv) participants were primarily identified and approached online, in many cases via social media. Despite this, the extent to which social media was viewed as the primary site of conflict between extremist groups and those opposed to them was striking. This was evident in the language used by many of the interviewees (e.g., inbox, viral, like, share, retweet, ban, playlist, etc.), suggesting that social media tools have become important lenses through which informal counter messaging actors view their practices and measure their achievements. Control of the social media narrative is, for many, equated to control of wider societal narratives and is therefore an end in itself. The focus on social media may go some way to explaining why those interviewed for this project often distanced themselves from those they saw as engaging in physical confrontation.

Lastly, it is entirely correct to be concerned about the use of online messages to spread or contest extremism. What societies do and say in response to extremist views matters greatly. This research has attempted to go a level below the schemes, plans, and programs that are typically featured in the literature. While the informal actors featured here are in some ways freer to be more direct and to exploit their own cultural and social resources, there is a more integral level still. Every day, millions of conversations take place between friends, colleagues and family members which serve to elevate the social costs of extremist activism. To be a member of an extremist group is, to some extent, to cut yourself off from mainstream political and religious society. While taking into consideration the role of informal actors in opposing extremism is a worthwhile activity, more also needs to be done to understand the roles of friends and family.

Benjamin Lee, Department of Politics Philosophy and Religion, Lancaster University, Lancaster, United Kingdom. [b.lee10@lancaster.ac.uk].

Notes

This work was funded by the Centre for Research and Evidence on Security Threats (CREST). CREST is commissioned by the Economic and Social Research Council (ESRC Award: ES/N009614/1) with funding from the UK Intelligence Community.

1. Drawing a distinction between formal and informal actors is difficult and necessarily imprecise as it ultimately turns on the perceptions of audiences that are not part of this study. In this study, informal actors are defined by (i) not being employed by any organisation linked to counter messaging, and (ii) operating independently or as part of a small group. This was combined with the researcher's own judgements about how he perceived the actor. In one instance an interviewee was actively employed in the CVE sector, but also undertook a substantial amount of counter messaging independently.
2. See: <https://www.youtube.com/watch?v=AIPD8qHhtVU>.
3. See: <https://www.reddit.com/r/beholdthemasterrace/>.
4. See: https://twitter.com/Mock_The_Right.
5. See: <https://redirectmethod.org/>.
6. See: <http://www.joindisruptors.com/>.

References

- Aistrophe, T. 2016. "The Muslim Paranoia Narrative in Counter-Radicalisation Policy." *Critical Studies on Terrorism* 9: 182–204. <https://doi.org/10.1080/17539153.2016.1175272>
- Bartlett, J., and A. Krasodomski-Jones. 2015. *Counter-Speech: Examining Content That Challenges Extremism Online*. London: Demos.
- Braddock, K., and J. Horgan. 2016. "Towards a Guide for Constructing and Disseminating Counter-Narratives to Reduce Support for Terrorism." *Studies in Conflict & Terrorism* 39 (5): 381–404. <https://doi.org/10.1080/1057610X.2015.1116277>
- Braddock, K., and J.F. Morrison. 2018. "Cultivating Trust and Perceptions of Source Credibility in Online Counternarratives Intended to Reduce Support for Terrorism." *Studies in Conflict & Terrorism*, 1–2. <https://doi.org/10.1080/1057610X.2018.1452728>, <https://www.tandfonline.com/doi/full/10.1080/1057610X.2018.1452728>. Advance online publication.

- Briggs, R., and S. Feve. 2013. *Review of Programs to Counter Narratives of Violent Extremism: What Works and What are the Implications for Government?* London: Institute For Strategic Dialogue.
- Cobain, I., A. Ross, R. Evans, and M. Mahmood. 2016. "Help for Syria: The 'Aid Campaign' Secretly Run by the UK Government," <https://www.theguardian.com/world/2016/may/03/help-for-syria-aid-campaign-secretly-run-by-uk-government>.
- CREST. 2018. "Introductory Guide: Countering Violent Extremism." Lancaster: Centre For Research and Evidence on Security Threats. <https://crestresearch.ac.uk/resources/countering-violent-extremism/>.
- Dupont, B. 2004. "Security in the Age of Networks." *Policing and Society* 14 (1): 76–91. <https://doi.org/10.1080/1043946042000181575>
- Eatwell, R. 2006. "Community Cohesion and Cumulative Extremism in Contemporary Britain." *The Political Quarterly* 77 (2): 204–16. <https://doi.org/10.1111/j.1467-923X.2006.00763.x>
- Ernst, J., J.B. Schmitt, D. Rieger, A.K. Beier, P. Vorderer, G. Bente, and H.-J. Roth. 2017. "Hate Beneath the Counter Speech? A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos." *Journal for Deradicalization* 10: 1–49.
- Ferguson, K. 2016. "Countering Violent Extremism Through Media and Communication Strategies: A Review of the Evidence." <http://www.paccsresearch.org.uk/wp-content/uploads/2016/03/Countering-Violent-Extremism-Through-Media-and-Communication-Strategies-.pdf>.
- Frenett, R., and Dow, M. (2015). *One to One Interventions: A Pilot CVE Methodology*. London: Curtin University and Institute for Strategic Dialogue.
- Glazzard, A. 2017. *Losing the Plot: Narrative, Counter-Narrative and Violent Extremism*, The Hague: International Centre for Counter Terrorism.
- Goodall, H.L., P.H. Cheong, K. Fleischer, and S.R. Corman. 2012. "Rhetorical Charms : The Promise and Pitfalls of Humor and Ridicule as Strategies to Counter Extremist Narratives." *Perspectives on Terrorism* 6 (1): 70–9.
- Harris-Hogan, S., K. Barrelle, and A. Zammit. 2016. "What is Countering Violent Extremism? Exploring CVE Policy and Practice in Australia." *Behavioral Sciences of Terrorism and Political Aggression* 8 (1): 6–24. <https://doi.org/10.1080/19434472.2015.1104710>
- Huey, L., J. Nhan, and R. Broll. 2013. "'Uppity Civilians' and 'Cyber-Vigilantes': The Role of the General Public in Policing Cyber-Crime." *Criminology and Criminal Justice* 13 (1): 81–97. <https://doi.org/10.1177/1748895812448086>
- Lee, B. 2016. "'It's Not Paranoia When They Are Really Out to Get You': The Role of Conspiracy Theories in the Context of Heightened Security." *Behavioral Sciences of Terrorism and Political Aggression* 9 (1): 4–20.
- Lee, B. 2019. "Informal Counter Messaging: The Potential and Perils of the Informal Counter Messaging Space." *Studies in Conflict and Terrorism* 42 (1–2): 161–77.
- Neumann, P.R. 2013. "Options and Strategies for Countering Online Radicalization in the United States." *Studies in Conflict & Terrorism* 36 (6): 431–59. <https://doi.org/10.1080/1057610X.2013.784568>
- Nhan, J., L. Huey, and R. Broll. 2017. "Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings." *British Journal of Criminology* 57 (2): 341–361. <https://doi.org/10.1093/bjc/azv118>.
- Payne, K. 2009. "Winning the Battle of Ideas: Propaganda, Ideology, and Terror." *Studies in Conflict & Terrorism* 32 (2): 109–28. <https://doi.org/10.1080/10576100802627738>
- Schmid, A.P. 2014. *Al-Qaeda's 'Single Narrative' and Attempts to Develop Counter-Narratives: The State of Knowledge*, The Hague: International Centre for Counter Terrorism.
- Sedgwick, M. 2012. "Jihadist Ideology, Western Counter-Ideology, and the ABC Model." *Critical Studies on Terrorism* 5 (3): 359–372. <https://doi.org/10.1080/17539153.2012.723520>.
- Silverman, T., C. Stewart, Z. Amanullah, and J. Birdwell. 2016. *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.

- Tierney, M. 2017. "Combating Homegrown Extremism: Assessing Common Critiques and New Approaches for CVE in North America." *Journal of Policing, Intelligence and Counter Terrorism* 12 (1): 66–73. <https://doi.org/10.1080/18335330.2017.1293281>
- True, J, and Eddyono, S. 2018). *Preventing Violent Extremism: Gender Perspectives and Women's Roles*. Melbourne, Australia: Monash University. https://monash.figshare.com/articles/Preventing_Conflict_and_Countering_Violent_Extremism_through_Women_s_Empowerment_and_Civil_Society_Mobilisation_in_Indonesia/7100873