ONE TO ONE ONLINE INTERVENTIONS

A PILOT CVE METHODOLOGY

Curtin University

Institute *for* Strategic Dialogue

About the Authors

Ross Frenett is a Fellow at the Institute for Strategic Dialogue and the Founding Director of Moonshot CVE, a specialist CVE focused organisation which aims to develop emerging methodologies to counter violent extremism. Ross previously served as Director of the Against Violent Extremism (AVE) network, a global network of former extremists and survivors of violent extremism seeded by Google Ideas and managed by the Institute for Strategic Dialogue. Throughout his career Ross has interviewed hundreds of former members of extremist groups and is a regular media commentator around extremism. He holds a Masters in Terrorism Security and Society from Kings College London and a BA from University College Cork

Moli Dow joined the Institute for Strategic Dialogue in 2014 as a Programme Associate to help manage, maintain and grow online counter-extremism initiatives. Working on a variety of online Countering Violent Extremism (CVE) networks, such as the Against Violent Extremism initiative (AVE), One2One and YouthCAN, Moli's experience ranges from both upstream engagement with civil society to downstream targeted messaging and online intervention for countering violent extremism. Having previously worked as a researcher at the Richardson Institute at the University of Lancaster, Moli's expertise includes both qualitative and quantitative methods to give an interdisciplinary approach to the initiatives she engages with. Moli holds a First Class degree in International Relations and Politics from the University of Lancaster.

About ISD

The Institute for Strategic Dialogue (ISD) is not a standard think tank, we aim to "do" as much as we "think." We build networks of credible messengers and partner with the private sector to ensure that we target and scale effective counter-narrative projects sustainably, and sensitively. These counter-narratives are essential to draining the extremist swamp of new recruits. We use a data-driven approach with support from leading social media companies such as Google, Twitter, and Facebook. By turning analysis into action, and 'learning by doing', ISD is at the forefront of a new approach to countering extremism that breaks free of familiar cycles of only reinforcing assumptions about extremism without demonstrating what works to counter it.

About Curtin University

This project was made possible by the support of **Dr Anne Aly** , Associate Professor, Curtin University: Countering Online Violent Extremism Research Program.The Countering Online Violent Extremism Research (COVER) Program is a hub of multi-disciplinary research activity dedicated to understanding the phenomenon of radicalization – online and in other forms. COVER have a focus on social media as a tool for radicalisation and countering violent extremism (CVE). They also research in the area of civil society movements and their roles in CVE.

INTRODUCTION

“Direct Intervention with those at Risk”

This report outlines the results of a pilot project undertaken by the Institute for Strategic Dialogue (ISD) in partnership with Curtin University and members of the global Against Violent Extremism (AVE) network.

This project was carried out in order to assess the viability of a methodology for direct intervention with those at risk of falling into the orbit of violent extremist organisations.

This report details the results of the pilot programme, outlines lessons learned and invites other organisations to critique and replicate these results.

BACKGROUND

The internet permeates all aspects of modern life and violent extremism is no exception. Although the level of importance is sometimes disputed, few would deny the role that online communication has in driving people towards violent extremist groups. While there are various perspectives on the exact nature of this process, it is increasingly agreed upon that it is rare for individuals to radicalise entirely in absence of any outside communication. Radicalisation remains a social phenomenon and the fact that some of these social interactions have migrated online does not change this. Extremists do not simply produce and disseminate propaganda and then move straight to offline recruitment, they utilise peer to peer messaging applications contained within social media platforms to engage in direct personal contact with potential recruits to their cause. Sometimes these online conversations completely displace offline recruitment.

Extremist propaganda is often removed and there are examples of nascent efforts to counter this throught the creation of counter-narrative campaigns. Counter-narratives, and offline counter-recruitment programmes such as EXIT and Channel, counter efforts of extremists to promote propaganda online and recruit in the offline world. This highlights the fact that there is a crucial piece missing in our efforts to counter recruitment to extremist groups; the proactive utilisation of peer to peer messaging systems online to engage with those expressing extremist sympathies.

Over the course of ISD's management of the AVE network we encountered a number of isolated attempts to engage directly with extremists online, from former extremists infiltrating extremist forums to Twitter conversations between activists and extremist sympathisers. However none of these efforts had been attempted at scale none had had testing and success metrics built in from the start. As such, any evidence gained as to their effectiveness was anecdotal at best.

PROJECT AIMS

This project aimed to test the viability of an approach based on directly messaging those openly expressing extremist sentiment online and seeking to dissuade them from following that path. In doing so the project team designed a methodology for identifying candidates, overcoming the technological barriers, and measuring which messages were most effective in eliciting responses most likely to lead to longer-term engagement.

METHODOLOGY

Data such as age and location, was key to matching a candidate with the appropriate intervention provider

As the aim of this project was to test a methodology which could be replicated and utilised by other organisations it was decided to avoid costly bespoke software to identify and engage those expressing extremist ideas online and instead rely on tools freely available to all. Similarly, the pilot aimed to test this methodology across ideologies and geographies.

This section will detail the chosen methodology and the rationale underpinning it.

PLATFORM

Although other platforms were explored, it was decided that Facebook was the most appropriate site to test this methodology for a variety of reasons, as detailed below.

CANDIDATE IDENTIFICATION

Key to the success of this methodology is the ability to identify candidates that are at risk of falling into the orbit of extremist groups, or are already expressing sympathy for such groups. Although the structure of Twitter allows researchers relatively easy access to large amounts of data, Twitter users share far less personal data on their accounts. In the One to One project this data, such as age and location, was key to matching a candidate with the apocopate intervention provider.

The pilot project was in the design phase as Facebook Graph Search was being rolled out across the platform.Graph Search (as it operated at the time) allowed users to search Facebook by demographic factors such as age, location and relationship status, in addition to interests, pages liked and group membership. Therefore while Twitter is often the go-to platform for researchers seeking to map extremist ecosystems, Facebook was deemed to be more appropriate when attempting to identify at risk individuals within certain demographic criteria and matching those individuals to intervention providers.

MESSAGING SYSTEM

Facebook has an inbuilt and highly popular messaging system which allows accounts to communicate directly and privately with one another. This system has been used by a number of extremist recruiters to speak with potential recruits. Facebook users must be friends (or friends of friends) in order to message one another, with messages from unconnected accounts generally landing in an inbox marked 'Other'. This, for all intents and purposes, acts as a spam folder and is rarely checked.

However, Facebook does contain a little known piece of functionality called Pay To Message that, as the name indicates, allows users to pay a nominal fee to ensure that they can message accounts they are not connected to. This fee is typically as low as $1, although this does appear to fluctuate depending on the popularity of the individual involved. Once the fee is paid, Facebook messenger functions in the standard fashion, with read receipts that allow users to tell when their messages have been seen by the user they are messaging. Unlike Twitter, Facebook does not limit the number of characters which can be used in message. This allows for far more in depth initial messages.

Therefore, Graph Search allowed researchers to identify candidates, Pay to Message allowed outreach providers to reach those candidates and the sheer scale of Facebook meant that if we succeeded in developing a robust intervention methodology on this platform, it would likely be scalable to differing contexts.

INTERVENTION PROVIDERS

As this project aimed, in part, to bring lessons learned in offline interventions to the online world it was decided that the intervention providers selected to work on this project should all have experience carrying out offline interventions. Working on the assumption that former extremists would be the most credible messengers, the project team identified ten former extremists from within the AVE network - five former far-right extremists from North America and five former Islamist extremists from the UK. There was one woman and four men in each ideological grouping. This group allowed us to test the methodology across geographies, ideological grounds and gender. These intervention providers were offered a nominal honorarium of £100 a month, and were therefore acting on an almost entirely voluntary basis.

Intervention providers were given the option of remaining anonymous and all providers were allocated new Facebook accounts set up by the project team for the purpose if this project. These accounts were monitored by the project team throughout in order to code messages and responses, ensure a prompt reaction to responsive candidates and safeguard both intervention providers and candidates should a threat of imminent violence or recruitment arise.

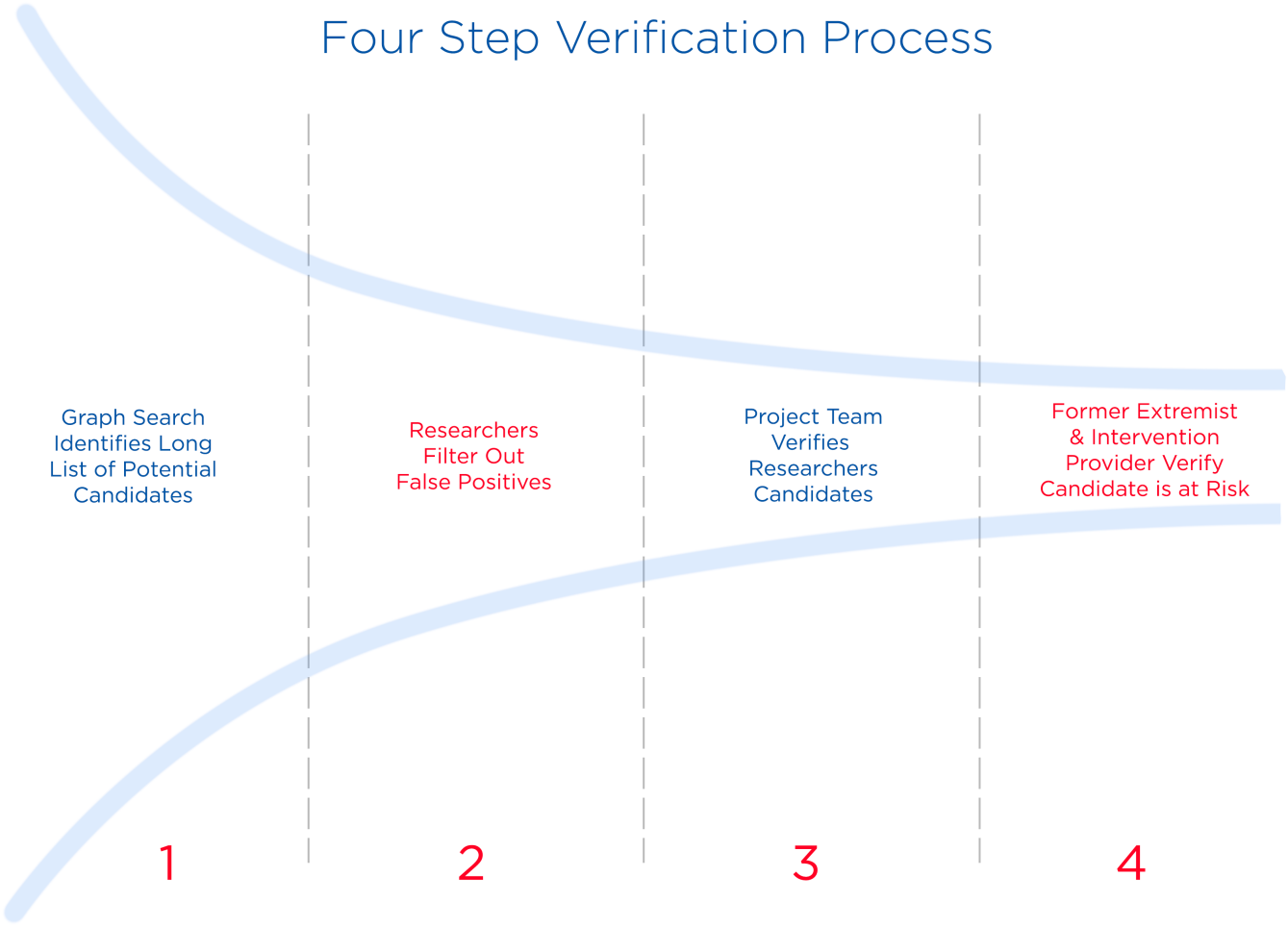
TARGETING CRITERIA & CANDIDATE INFORMATION

Once experienced intervention providers were identified and had agreed to participate, the targeting criteria which would dictate who would be considered 'at risk' had to be agreed upon. In order to achieve this, the project team built a list of the numerous factors Facebook Graph Search allowed users to search by, and circulated these among intervention providers. Although many of these were likely irrelevant ('device type' for example) it was decided to draw the first draft of the risk criteria from the intervention providers without steer, rather than have the project team dictate these.

The most relevant factors found to indicate risk were Pages liked and Group membership. There were also additional markers, such as a user's cover photo and the tone and content of their regular posts. The project team collated the risk factors provided by the outreach providers and created two distinct sets of search criteria -one for those considered at risk of falling into the orbit of the violent Far Right extremism in north America and another for those considered at risk of falling into violent Islamism in the UK.

Having signed non-disclosure agreements and undertaken training, external researchers recruited for the duration of this project began to utilise Graph Search to identify candidates they believed to be at risk. To ensure the candidates were genuinely at risk, the candidates were reviewed first by the project team and then by the intervention providers. Two categories of people were removed at this stage -those that had been falsely identified as being at risk, as they were following pages and accounts which were politically or religiously extreme but did not advocate violence, or those that were so far into the extremist milieu they went beyond the scope of this project. One example where this was evident saw an intervention provider declining to contact their candidate as they considered theircontacting this person would put their personal safety at risk.

The mercurial nature of the extremist ecosystem on Facebook meant that the original targeting criteria acted as a starting point and evolved throughout the life of the project, with certain pages and groups taken down and others, that were found to be common among those at risk added on.



MESSAGE TYPE

As this was a pilot project there was no established best practice as to what messages were most likely to resonate with such a unique target audience. In offline interventions there are competing theories as to what tone is most appropriate and whether or not to engage in ideological discussions. As such, it was decided the project team would not provide any direction to intervention providers as to what message to send but. This allowed each provider to build unique messages based on their own intervention experience and personal background. This had the added benefit of allowing the project team to gather data on what message types were most likely to elicit a response, and add much needed data to an on-going discussion between intervention providers of all kinds.

RESULTS

In the UK a number of those candidates deemed to be at risk of imminent travel to Syria were referred to Police contacts within the Channel programme

Data was gathered on an on-going basis throughout this project in order to provide a strong foundation for future projects interested in utilising this methodology.

Some of the key findings are detailed below, but these results are by no means exhaustive.

PROFILE IDENTIFICATION & VERIFICATION

The project aimed to identify 160 profiles of those considered to be at risk of carrying out violence, drawing 80 from the violent Far Right in North America and 80 from the violent Islamist movement in the UK. In total 154 profiles were identified. The discrepancy between the target and the actual number identified was due to the inability of one of the intervention providers to complete his work. As this withdrawal took place before we had identified all candidates for this provider, it was decided researchers would not identify profiles that could not be acted upon. As outlined above, once identified profiles were passed to intervention providers to verify, over 90% of the profiles identified were confirmed as being 'at risk'.

6%	Insufficient Risk
1%	Severe Risk
93%	At Risk

In the UK, a number of the candidates deemed to be at risk of imminent travel to Syria were referred to police contacts within the Channel programme, a UK Government initiative which aims to provide support to individuals at risk of being drawn into violent extremism[1].

Word Cloud of all pages 'Liked' by those at risk of Falling into the Orbit of Violent Extremists

1 *For more information, see here: https://www.counterextremism.org/resources/details/id/115/channel-process*

COMMON PAGES AMONG THOSE AT RISK OF FALLING INTO THE ORBIT OF VIOLENT ISLAMIST GROUPS

Of those confirmed to be at risk of falling into the orbit of violent Islamist groups there were a number of pages which a high proportion of candidates were connected to. These are detailed in the below table. It is important to note that while some of these pages would indicate risk (for example those openly associating themselves with Islamic State of Iraq and the Levant), others are more broadly political or religious in nature and would not necessarily act an indicator in isolation.

Group Name	Percentage of those at Risk Connected
Authentic Tawheed	17.57%
Yusha Evans	13.51%
Prisoners of Faith	12.16%
Tawheed Network	12.16%
Anwar al-Awlaki	8.11%
Women in West for Khilafah	8.11%
Anjem Choudry	6.76%
Anwar Awlaki	6.76%
Islamic State of Iraq and the Levant	6.76%
Abdulrahman Muhajir	5.41%
AlKhilafah	5.41%
Green Birds	5.41%
Lions of Tawheed	5.41%
Ummah News	5.41%
Authentic Tawheed	5.41%

COMMON PAGES AMONG THOSE AT RISK OF FALLING INTO THE ORBIT OF VIOLENT FAR RIGHT GROUPS

Of those confirmed to be at risk of falling into the orbit of violent Far Right groups there were a number of pages which a high proportion of candidates had connected to. These are detailed in the below table. As evidenced, the Far Right candidates were a more heterogeneous group than those at risk of falling into the orbit of violent Islamist groups, with only one page present among more than 6% of candidates. It is important to note that while some of these pages would indicate risk, others are more broadly political (for example the Tea Party) or religious in nature would not necessarily act an indicator in isolation.

Group Name	Percentage of those at Risk Connected
Smash Cultural Marxism	7.5%
Aryan Nations	6.3%
The White Voice Network	6.3%
Aryan Brotherhood	5%
Right Wing News	5%
Justice for Germans	3.8%
Klu Klux Klan	3.8%
National Front	3.8%
Stop White Genocide	3.8%
The Tea Party	3.8%
White Aryan Resistance	3.8%
White People	3.8%
14/88	2.5%
Anti-Communism	2.5%
National Alliance of White European Americans	2.5%
National Socialism	2.5%

Word Cloud of all pages 'Liked' by those at risk of Falling into the Orbit of Violent Far Right Groups

MESSAGES SENT AND SEEN

Of the profiles confirmed to be at risk, 76 messages were drafted and sent to candidates. The large difference between the number of profiles identified and the number of messages sent was due to a number of factors, both technical and human, including; the removal of profiles from Facebook, the withdrawal of designated intervention providers, malfunctioning pay to message functionality and of course those profiles deemed to be of insufficient risk. Over 60% of the messages which were sent were actually seen by the candidate, with no notable difference between Far Right and Islamist profiles. This read rate is significantly higher than standard unsolicited mailing campaigns. According to MailChimp -a specialist email marketing service provider -no industry has an average read rate which exceeds 30%[2].

> *"Over 60% of the messages which were sent were actually seen by the candidate"*

2 *See Mailchimp's email marketing benchmarks here: http://mailchimp.com/resources/research/email-marketing-benchmarks/ (Last accessed 01/09/2015)*

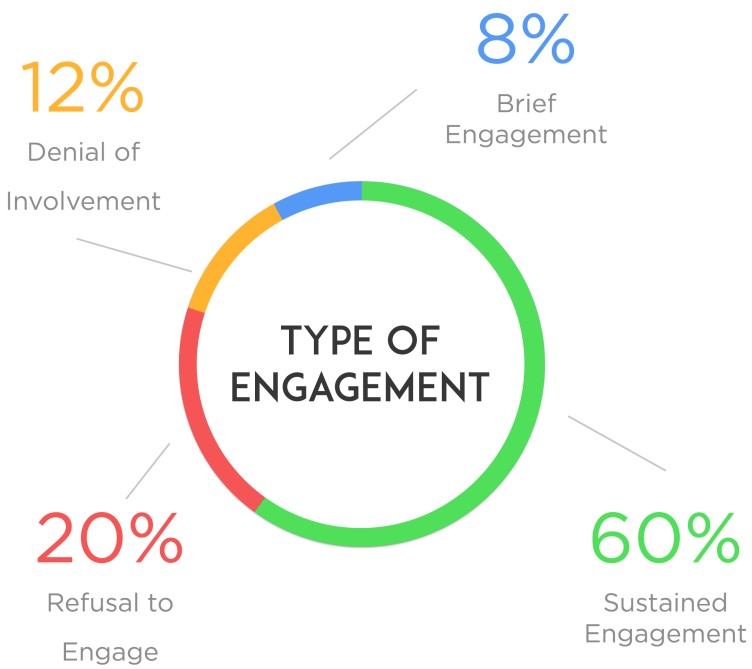
CANDIDATES REACTIONS

Candidates that were successfully reached (i.e. those that saw a message sent to them) were highly likely to react in some way, with 59% responding directly or demonstrating a shift in behaviour, such as a change in privacy settings or blocking the provider. These shifts in behaviour, in absence of an actual reply, were coded as 'reactions'. A majority of candidates sent a response of some kind; however there was a notable difference in response rates of far-right and Jihadist candidates,with 63% of far-right candidates responding as opposed to 42% of Islamists. Again, comparing this reaction rate with standard email marketing campaigns, this is orders of magnitude higher than would be expected. No industry tracked by MailChimp achieves click rates (indicating the person receiving the email is interacting with the content) higher than 6%.

One possible reason for the differing response rates between groups is the legal frameworks in the UK and the United States. In the US candidates openly express support for a racist ideology and pose with firearms. This may suggest US candidates are more likely to believe they are immune from prosecution due to the first and second amendments, and as such more likely to engage. In addition, the media surrounding ISIS and associated groups during this pilot project may have encouraged those candidates expressing sympathy for ISIS to be more cautious.

RESPONSE TYPES

Of those candidates that did respond, a proportion either denied their adherence to the ideology in question or refused to engage. However a majority were willing to engage in a sustained conversation with an outreach provider. Sustained engagement was coded if five or more messages were exchanged and the candidate entered into a meaningful conversation with the intervention provider.



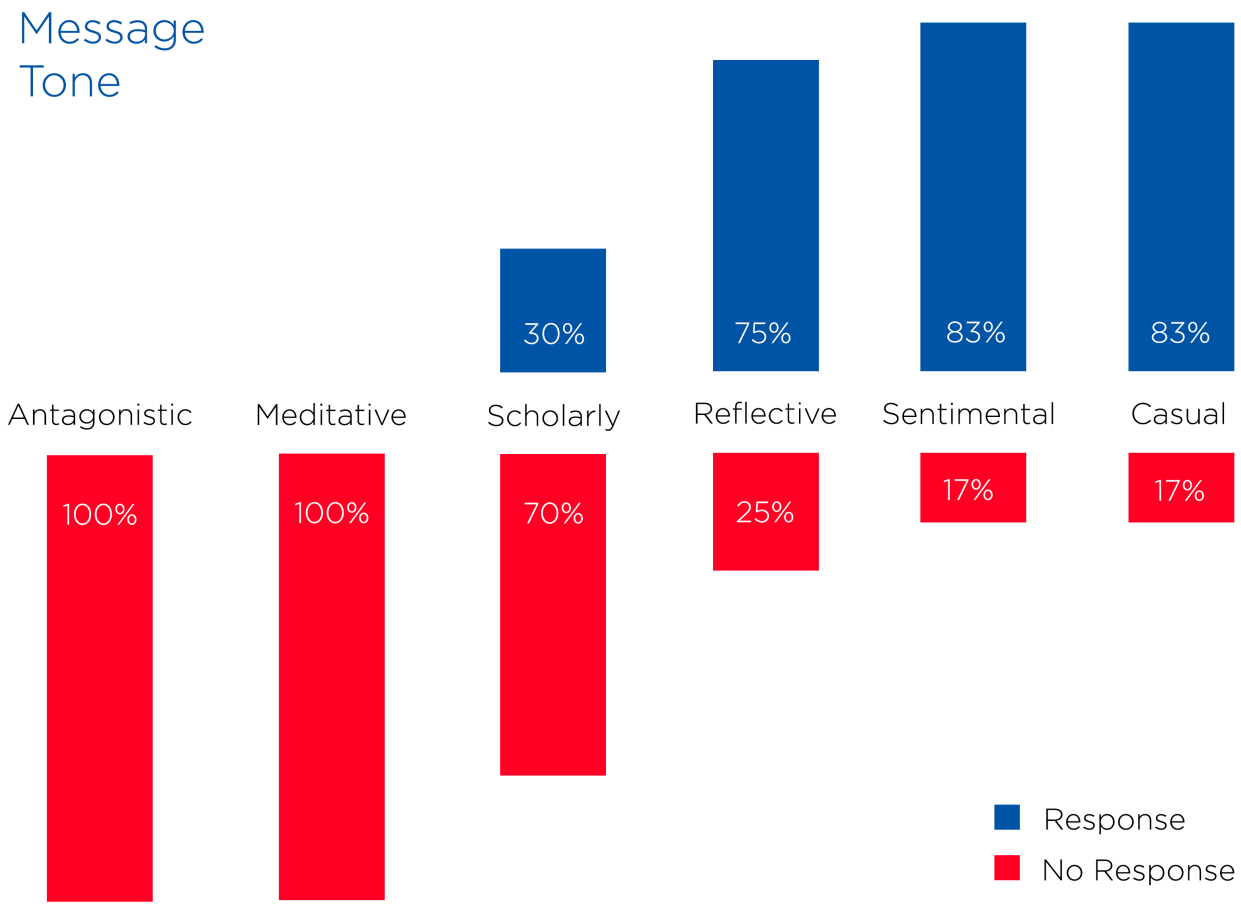
VARIABLES AFFECTING RESPONSE TYPE

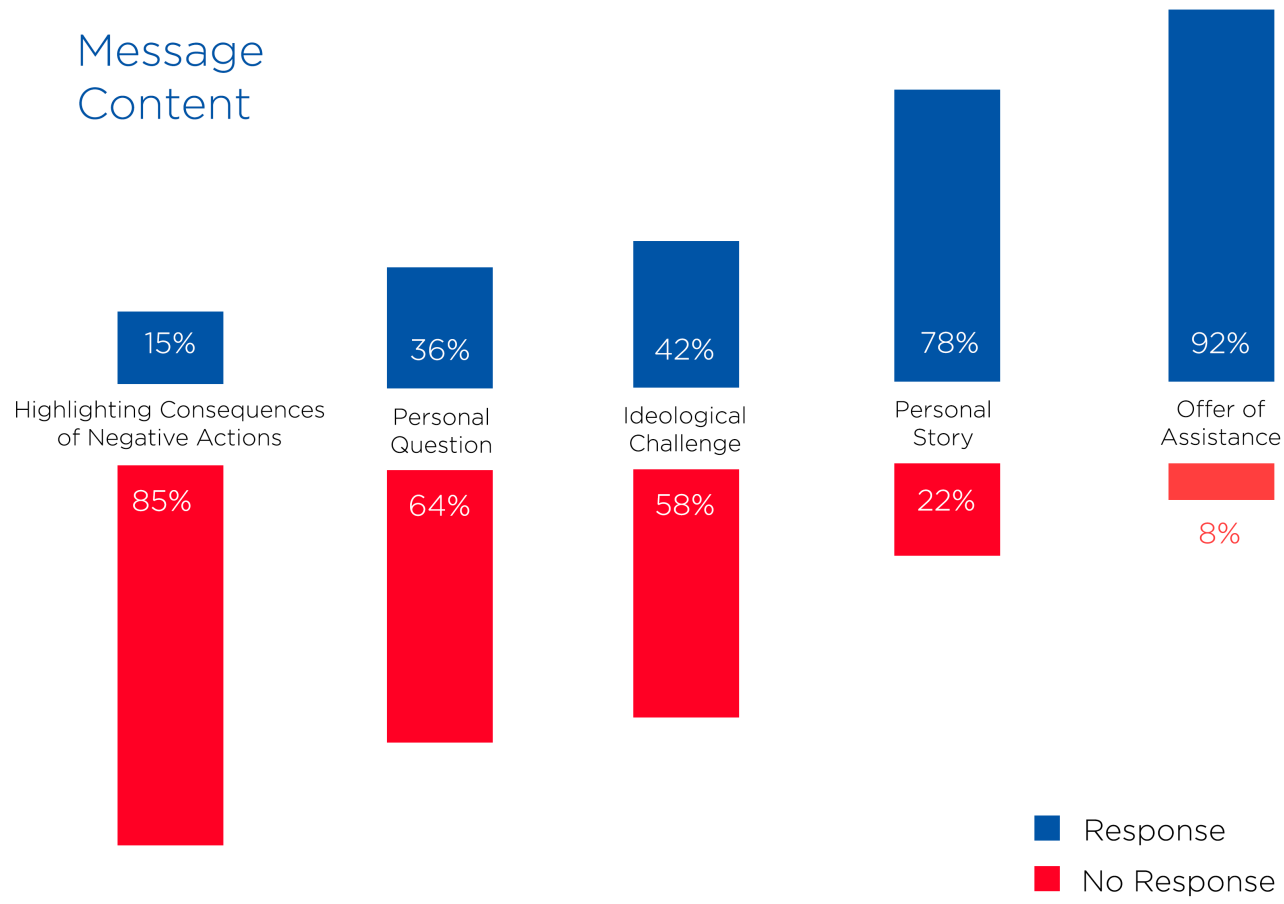
Response rates varied by large margins between providers; one provider received a response rate of over 90% and two received no responses at all. At first glance it would appear to confirm the suspicion of many offline providers that there is an unquantifiable 'x factor' that some intervention providers possess and others do not. However, when capturing the messages sent, the project team coded these by tone and content, identifying those factors which were common to the successful messages. This allowed us to start to build data-driven best practice for online intervention. These are detailed below.

TONE

Message tone is highly important, with certain tones eliciting zero responses and others, such as the casual and sentimental, eliciting over 80% response rates. A casual message, for example, could involve a simple prompt such as, "Hey, you're an MMA fighter?" Whilst, a sentimental message, included language such as, "I understand" and "I know what you must be feeling". A reflective message tended to involve intervention providers examining their past and its effect on their future.

This indicates that a simple adjustment in tone may drastically increase the probability an intervention provider will receive a response.





CONTENT

In addition to tone, message content was categorised and measured. Certain patterns and best practices can be identified via the below results. These demonstrate, for example, the importance of personal stories and the futility of a negative approach.

Highly personal messages which drew on the personal experiences of the intervention provider and offered non-judgmental assistance to those at risk were the most effective. This may justify the decision to carry out this initial pilot project utilising former extremists, though it is possible the personal stories of survivors of extremism would be equally compelling.

Messages which were more than one sentence but less than five elicited the most responses. This suggests a less time consuming approach could be more effective. As, although one line prompts were largely seen as ineffectual, messages two paragraphs long or more were also less productive. Instead, response peaked at messages approximately a paragraph long.

> *"Highly Personal messages ... were the most effective"*

ANONYMITY

Certain intervention providers opted to carry out the interventions using pseudonyms and anonymous profiles for their personal safety. Although it was expected that anonymous profiles would be less effective at eliciting responses, in fact these accounts received slightly higher levels of engagement than those using real names. However the choice to remain anonymous did seem to have an effect on intervention providers' ability to achieve sustained engagement.

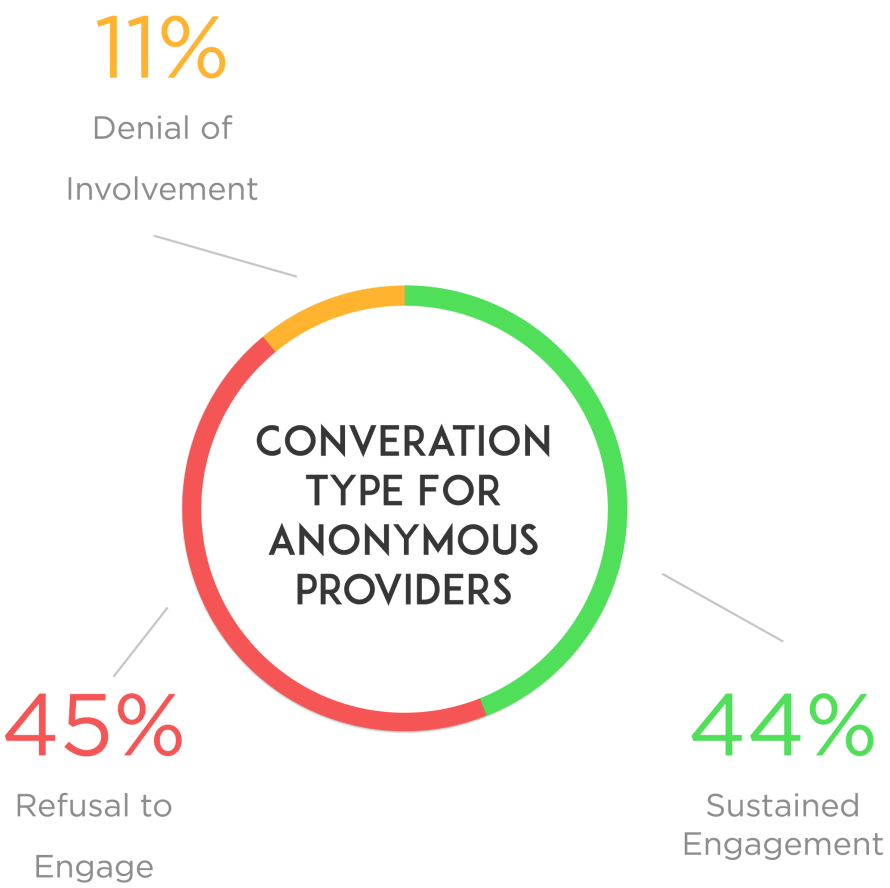
> *"These messages do not, contrary to initial expectations, have to be bespoke. This result will help to make the project more replicable and less time consuming for the outreach providers"*

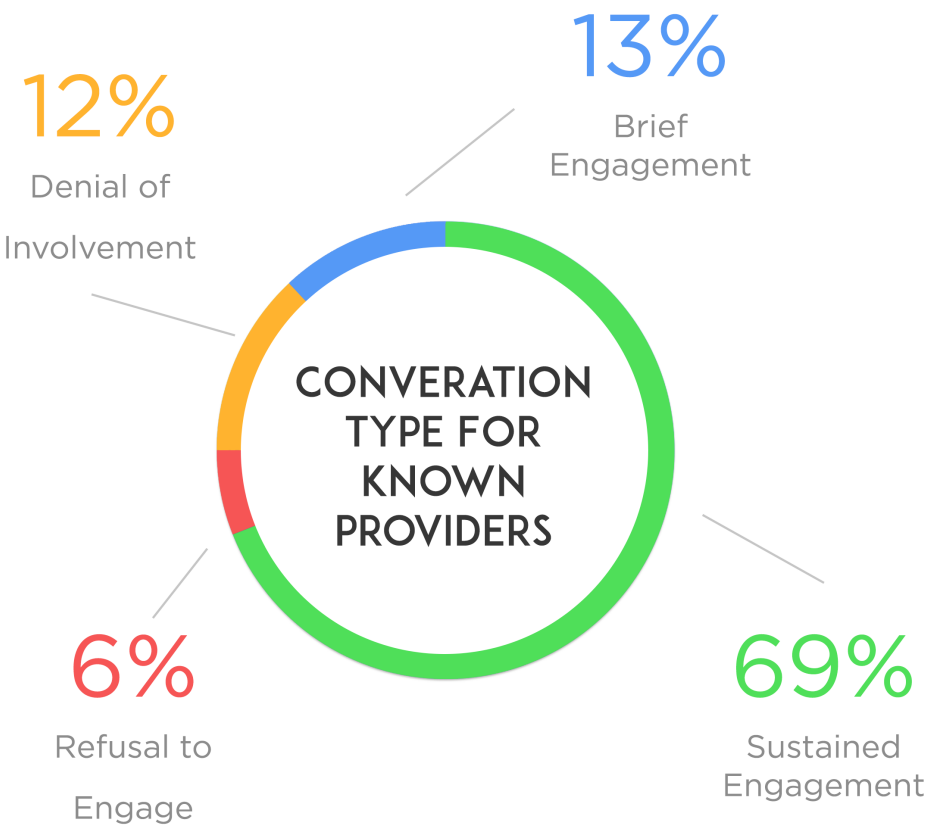
Certain candidates indicated that they would not engage with someone using a pseudonym, with one candidate responding to an anonymous intervention provider: "I can't take anything you say because you could be anyone. You are hiding behind a name." Another candidate responded to a message from a known account, stating "I read about you online after receiving this message" and "your life story is definitely different". These messages suggest that candidates may search online for the people contacting them. It is important to note, however, that this could also have a negative impact. One candidate responded to outreach attempts with "Get lost you **** foundation monafique!" This is another element that would benefit from further research, with a larger dataset.

SHIFT IN BEHAVIOUR

This project took place over too short a time period to effectively measure any long term shifts in belief system or behaviour. However a number of interactions gave indications that, with sustained engagement, achieving a long-term adjustment in behaviour may be possible, with trust built between intervention providers and candidates. This was evident when certain candidates stated they felt they could trust their outreach provider and that the interaction led them "to think deeply" for the first time.

Other interactions indicated a willingness to explore alternative ideas and engage with the personal journey of the intervention providers. For example, one candidate messaged"[I] am curious to hear how and why you've changed your mindset." Another intervention began with the candidate refusing help, stating "I don't need no help I like me for me." It was decided not to pressure those that refused, so the intervention was left at that. However, one week later the same candidate reached out to the intervention provider saying "I need help if u can help me." This highlights both the importance of establishing lines of communication with those at risk and the potential impact this methodology could have in the longer term .


11%
Denial of Involvement
CONVERATION TYPE FOR ANONYMOUS PROVIDERS
45%
Refusal to Engage
44%
Sustained Engagement


13%
Brief Engagement
12%
Denial of Involvement
CONVERATION TYPE FOR KNOWN PROVIDERS
6%
Refusal to Engage
69%
Sustained Engagement

LESSONS LEARNT

The experiences and expertise of the intervention providers were fundamental to the success of this project

The pilot project was an experiment to test the feasibility of outreach to at risk individuals online. As with all experiments, the learning curve was steep and many lessons have been learned throughout.

This section will explore the key lessons learned and seek to identify some of the pitfalls for other organisations to avoid.

TECHNOLOGY

Technology was essential to the project. However there was often a gap between what was theoretically possible with a piece of functionality on a platform and the reality. This project encountered several obstructions and delays as the technology was often unreliable and unpredictable.

GRAPH SEARCH

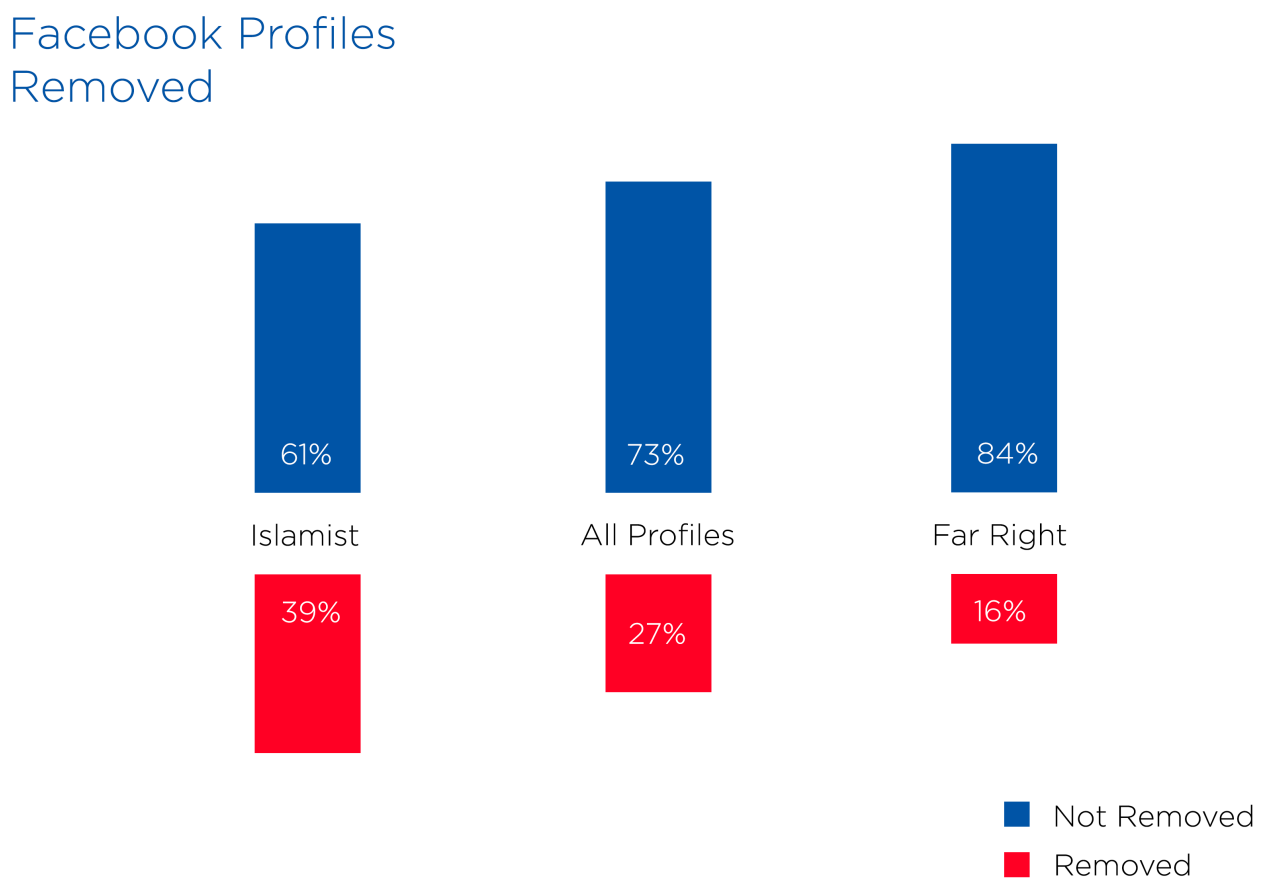
As outlined in the methodology section, Facebook was chosen as it allowed researchers to identify individuals that openly endorsed and promoted extremist narratives online. However, during the project, Graph Search became increasingly restricted, with both the search function and the results limited. In the later stages of the project this greatly slowed the rate at which candidate accounts could be identified.

ACCOUNT REMOVAL

The project was affected by the number of profiles which were removed by Facebook. As evidenced below, of the 154 profiles identified, 42 were removed by Facebook throughout the project. This issue disproportionately affected Islamist, rather than Far Right profiles. These difficulties affected the researchers' ability to collect and identify profiles.

PAY TO MESSAGE

Facebook was also selected as it supplied intervention providers with the technology to reach their candidates via the Pay to Message functionality. However, this was also found to be unpredictable. Throughout this pilot project, 'pay to message'



was seen to sporadically malfunction in all Facebook accounts. For two intervention providers, pay to message was permanently disabled, for no coherent reason that the project team could identify. In the absence of 'pay to message', candidates were largely unaware they were ever contacted.

> *"Overall, whilst these technological difficulties did not derail the project, they did delay it and may pose significant barriers to future attempts to utilise this methodology"*

INTERVENTION PROVIDER ACCOUNTS

Further difficulties regarding intervention providers' accounts were encountered. Facebook limits one individual to one Facebook account. Therefore, as intervention providers had two accounts in their name (one for their own personal use and another for the project), one intervention provider had his account frozen. These technical problems restricted intervention provider's attempts to connect and converse with candidates and, overall, whilst these technological difficulties did not derail the project, they did delay it and may pose significant barriers to future attempts to utilise this methodology.

INTERVENTION PROVIDERS

The experiences and expertise of the intervention providers were fundamental to the success of this project. However a number of serious issues were encountered throughout the project which could threaten the viability of attempts to replicate this methodology elsewhere.

The project plan required each intervention provider send one message per week for sixteen weeks. All providers had full-time jobs and were also involved in several other CVE projects, which could be time consuming and extremely draining. In addition, the project did not have the funds available to pay the intervention providers for their work, beyond a £100 monthly honorarium. Due to this throughout the project intervention providers struggled to adhere to their targets, and on a number of occasions had be to chased more than once to engage with responsive candidates. Four intervention providers had to be replaced due to their personal time constraints. Consequently - while waiting in their intervention providers queue - some Facebook profiles were taken down and, therefore, could no longer be contacted.

Another issue which emerged involved intervention providers' concerns over personal safety and security. As with all intervention work, there was an element of risk involved in contacting individuals that openly endorse and promote extremist content online. One intervention provider left the project due to their concerns, whilst three others chose to interact anonymously.

RECOMMENDATIONS

“NGOs that carry out intervention work, ... need to integrate online and offline intervention efforts”

This pilot highlighted the potential of this methodology to significantly alter how we counter radicalisation. However the dataset involved here is a small one; other organisations and researchers should test, replicate and improve on this methodology. In particular a number of the assumptions underpinning this pilot phase need to be tested, such as Facebook being the most appropriate medium and former extremists being the most effective messengers.

Platforms such as Twitter and Google+ should also be explored. The use of community leaders, survivors of violent extremism and specially trained online intervention providers should also be explored and data on their effectiveness gathered and shared.

The methodology employed during this pilot phase involved NGO actors utilising a private sector platform in consultation with Government agencies and programmes. As such, the key recommendations generated span all three sectors and will be divided out as such.

GOVERNMENT

- Online intervention work, much like offline intervention work, is a time consuming and difficult undertaking which is not sustainable unless intervention providers are highly motivated and well supported. Poorly resourced interventions which are sporadic risk exacerbating the problem rather than solving it. As such Governments need to fund interventions with at risk youth in the online world as well as offline.

- Government coordinated intervention programmes, such as Channel in the UK, need to be updated to streamline a process of online referrals, proactively seek those at risk online and break down the artificial barriers which exist between online and offline communication.

SOCIAL MEDIA COMPANIES

- Social media platforms such Facebook and Twitter should build on efforts piloted in other areas of user safety and grant approved NGOs enchased access to identify and assist vulnerable individuals. This could build upon a similar model to the suicide prevention tool which Facebook rolled out in early 2015, which allows users to flag posts of concern and either message the potentially suicidal person, contact another Facebook friend for support or connect with a trained professional at a suicide helpline for guidance[3].

- Processes should be established to ensure that accounts which are removed, or simply flagged, for promoting terrorist material are referred to independent NGOs that can assess the likelihood that the individual in question is a threat to themselves or others, and take appropriate action where necessary.

- A number of designated point people with swift access to technical knowledge required by NGOs and others working to assist vulnerable individuals should be appointed. With platforms that are in a contestant state of flux, having swift technical advice as to how to best navigate platforms would ameliorate many of the issues faced during this pilot phase, including security concerns.

INTERVENTION NGOS

- NGOs that carry out intervention work need to integrate online and offline intervention efforts, creating off roads for those at risk in the online space which can link in to their offline programmes.

3 *Newsbeat, "Facebook launches new suicide prevention tool in the US", BBC, 26th February 2015, link last accessed 29th July 2015, http://www.bbc.co.uk/newsbeat/article/31641216/facebook-launches-new-suicide-prevention-tool-in-the-us.*

CONCLUSION

Although peer to peer messaging has been a temporary boon to violent extremists online, it could be their ultimate downfall

For too long those countering extremism online have been on the defensive, with the tools the internet provides viewed as problems to be solved rather than opportunities to be exploited. However, as the multi-billion dollar companies which have sprung up to mine and exploit personal data illustrate, all internet users leave trails of data. These trails can be picked up , with the most at risk users identified and engaged with to steer them away from their dangerous path.

If honed and properly resourced this methodology points the way to a future in which an expression of violent extremist sentiment on social media is met in much the same way as it would be met offline - through the intervention of concerned specialists, rather than with recruitment or arrest. Although peer to peer messaging has been a temporary boon to violent extremists online, it could be their ultimate downfall.

APPENDIX

ETHICAL CONSIDERATIONS

The project involved an examination of individuals that openly endorsed and promoted extremist narratives online. In this process, both the quantitative and qualitative data of these individuals was also collected and analysed. Although this study was limited to publicly accessible social media data, the nature of Facebook as a platform means these individuals may have been under the impression that the opinions they expressed online would be private.

In relation to the collection and analysis of social media data, the project sought to follow the ethical guidelines laid out by both the AoIR Ethics Working Committee[4] and the British Psychological Society in 2013[5]. As advised in both guidelines, any research that entails amassing and examining social media data must balance the potential risk to those being studied, with the greater social good that the project could provide. This section will further examine and evaluate the balance between the risk undertaken in this project and the social good fostered by it.

RECOGNISING THE RISK

The pilot project entailed a degree of risk. The individuals examined within this project could have encountered issues concerning both their personal safety and mental wellbeing, particularly if their support of extremist content were to be exposed. Moreover, the risks were not limited to the individuals contacted in this project. The personal risk to the intervention providers - who contact and engage with the candidates - must also be recognised. However, the project's risk level must be measured against the social good produced by this work. Reaching out to vulnerable individuals will always entail an element of risk. Ultimately, these risks were deemed necessary to tackle such an extensive problem.

4 *British Psychological Society, Ethics Guidelines for Internet Mediated Research, Leicester, BPS (2013), Link last accessed December 3rd 2014. http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf.*

5 *AoIR, Ethical Decision Making and Internet Research: Recommendations from the AoIR Ethics Working Committee Version 2.0 , Chicago: AoIR (2012), Link last accessed December 3rd 2014. http://aoir.org/reports/ethics2.pdf.*

CONTROLS IN PLACE

Various precautions, which sought to ensure the safety of both the candidates and intervention providers, were taken in this project.

RISKS TO CANDIDATES

The risk to candidates revolved around any possible infringement on their privacy and the exposure of their identity. Therefore, to limit this risk, provisions were taken to safeguard the data sourced, analysed and captured during this project. To achieve this, the Data Protection Policy largely focused on securing both the electronic and hard records of the data.

- Information was only stored on an encrypted USB. When not in use, the USB stick was kept in a locked safe. Access to the safe was only available to ISD staff and the encrypted data could only be accessed by the project co-ordinators.
- All electronic documents stored at ISD were password protected and placed on the encrypted USB stick. Documents were only available on this USB stick.
- The USB and the documents within the USB never had the same password and these passwords were never communicated by email.
- Hard copies were only printed at ISD and could not leave the premises of ISD.
- Hard copies could not be kept for longer than two weeks and were stored overnight in a locked safe. At the end of the two week period these records were collected and sent to a secure firm for incineration.

RISKS TO INTERVENTION PROVIDERS

The risk to intervention providers primarily derived from the threats they would receive from the candidates, or any interested parties who discovered their work. The foremost concern was that this online risk would translate into an offline threat. It is also important to emphasise that all the intervention providers involved in this project are engaged in intervention work offline. Therefore, they are aware of the risks inherent in this work and have their own protection policies in place. Several precautions were taken to protect the personal safety of the project's intervention providers.

- The intervention providers used Facebook profiles created for the purposes of the project to perform outreach.
- The intervention provider's Facebook accounts were monitored and recorded by ISD co-ordinators. If any explicit threat of violent was evident, ISD co-ordinators would take direct action in reporting this to the correct authorities within an adequate time frame.
- Intervention providers were given the option to interact anonymously, if they wished to further protect their identity.



INSTITUTE *for*
STRATEGIC DIALOGUE