ISD
Powering new
generations
against extremism

COUNTER CONVERSATIONS

A model for direct engagement with individuals showing signs of radicalisation online

Jacob Davey

Jonathan Birdwell

Rebecca Skellett

About this paper

This report outlines the results of a programme trialling a methodology for identifying individuals who are demonstrating signs of radicalisation on social media, and engaging these individuals in direct, personalised and private 'counter-conversations' for the purpose of de-radicalisation from extremist ideology and disengagement from extremist movements. This is the first programme globally which has trialled the delivery of online interventions in a systematised and scaled fashion.

In no case may this material be altered, sold or rented. ISD does not generally take positions on policy issues. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the organisation.

Designed by forster.co.uk. Typeset by Lookandfeelstudio.

About the authors

Jacob Davey is a Researcher and Project Coordinator at ISD overseeing the development and delivery of a range of online counter-extremism initiatives including the counter conversations programme. His research interests include the role of communications technologies in intercommunal conflict, the use of internet culture in information operations, and the extreme right globally. He regularly provides commentary on counter-extremism issues and has provided expert advice to national and local policy makers.

Jonathan Birdwell is the Head of Policy and Research at ISD. Jonathan oversees ISD's Strong Cities Network, a global network of mayors, policy makers and practitioners working to build community resilience to violent extremism, as well as ISD's Policy Planners Network. Jonathan also oversees all of ISD's research, including the setting up and running of ISD's Digital Research unit. Jonathan's research interests include the relationship between violent and non-violent extremist groups, cumulative extremism, digital literacy, as well as broader questions around political participation and trust in institutions. Prior to joining ISD, Jonathan worked for seven years at the London-based think tank Demos where he published over 40 research reports including The New Face of Digital Populism and The Edge of Violence.

Rebecca Skellett is a Senior Programme Manager at ISD overseeing online and offline interventions. Through ISD's Strong Cities Network Rebecca has been involved in the delivery of training to over 2,500 municipal level practitioners, building capacity in recognising and responding to signs of radicalisation. Previously, Rebecca worked on the front-line of the UK's Prevent CVE Programme across several London boroughs, and has extensive experience in overseeing individual case work, conducting community engagement, and developing local CVE programming, training and policy frameworks for local government and the education sector.

Contents

Contents (cont.)

Executive Summary

Extremist groups deploy a clear strategy for radicalising and recruiting new supporters online: marketing their ideas through the spread of propaganda and then engaging interested individuals in direct, private messaging to recruit new members to their causes.

Direct engagement with radicalising individuals by mentors and 'intervention providers' is now a well-established component of offline counter-terrorism efforts in a number of countries. These programmes are delivered by both government and civil society, and often include former extremists and social workers as intervention providers.

Until now, online prevention efforts have largely focused either on the removal of terrorist content or on the production and dissemination of counter-narrative and counter-speech campaigns to compete with extremist propaganda. However, there have been no systematised attempts to supplement counter-speech efforts with direct online messaging and engagement at scale.

ISD's Counter Conversations programme is an experimental approach designed to fill this gap and test if the methods deployed in offline interventions can be brought into the social media domain. Delivered on Facebook to date and working across Extreme Right and Islamist ideologies, the programme provides an opportunity for individuals showing clear signs of radicalisation to meet and engage with someone that can support their exit from hate.

In this report, we present the findings of our most recent pilot programme of Counter Conversations. The results demonstrate the positive potential of direct online engagements and point to the need for further exploration into how this model can be deployed in a responsible, effective and scaled fashion, as part of a suite of online risk reduction methodologies.

Direct engagement with radicalised individuals

For the past three years, ISD has been working to appropriate the direct peer-to-peer messaging approach used by extremists and apply it to the online world.

This started with an initial small-scale pilot, conducted in 2015, using former extremists drawn from ISD's Against Violent Extremism (AVE) network to reach out online to over 150 Islamist and right wing extremists.[1] The results suggested that direct online outreach was a potentially viable tactic that was worthy of further exploration.

Accordingly, we developed a larger programme of action research designed to test the viability of delivering this work at scale in a secure and evidence-based fashion, and within a robust ethical framework.

With financial support from Facebook, this included:

- Developing a **semi-automated identification methodology** to efficiently yet carefully identify individuals publically supporting extremism and using violent language on social media platforms;

- Recruiting, training and testing the effectiveness of **different types of intervention providers**, including former extremists, survivors of extremist violence, and professional counsellors; and,

- Developing a r**obust risk assessment framework and safeguarding mechanisms** for undertaking direct outreach with individuals showing open signs of radicalisation online.

Identification

During the initial pilot programme in 2015, the identification of candidates for intervention was performed entirely on a manual basis by ISD experts and researchers. The first step in scaling direct engagement online was to develop a semi-automated identification methodology that carefully and accurately identifies individuals who are publically expressing signs of ideologically inspired hatred and violent sentiment towards others on social media.[2] This methodology consisted of:

1. Identification of Facebook accounts which were repeatedly engaging with public Facebook pages associated with the extreme right or Islamist extremism, or which tended to attract individuals expressing violently extreme viewpoints. In this fashion over 42,000 individuals were identified, an overwhelming majority of which were extreme right.

2. We then applied an approach that combined machine learning and a Natural Language Processing (NLP) algorithm to identify people who appeared to be using violent and dehumanising language against other groups of people on these pages.[3] This distilled the pool of individuals identified in step one to a group of around 7,000, of which a sample of 1,600 were analysed further.[4]

3. A process of manual review was then carried out by ISD experts and researchers on this sample of 1,600 individuals, gathering additional open-source information that indicated support for violent extremism including: profile pictures containing extremist imagery; likes / positive comments on pieces of content supporting violently extreme groups; likes / positive comments on extremist material; posts containing support for violently extreme groups; posts containing extremist material; friends within extremist networks; indication of offline involvement with extremist groups. This manual review also applied a risk-assessment framework based on well-established social work practices and purpose built for online interventions. Over 800 individuals were selected in this manner to be candidates for intervention.

Profile of Candidates for Engagement

Using open source data we were able to determine age, gender and other demographic details for individuals selected for online outreach. We found that:

- Islamist candidates were significantly younger than extreme right candidates: 3 out of 4 (72%) of extreme right candidates identified were over 45, while a slightly higher proportion (77%) of Islamist candidates identified were under 30.

- Female Islamist candidates presented the youngest age profile: 4 out of 5 (81%) were under 30.[5]

- Nearly 1 in 10 (9%) of male extreme right-wing candidates appeared to be current or former members of the armed forces

Intervention and Engagement

Just under three quarters of those candidates selected for online intervention (70%) were engaged by ISD's intervention providers, who initiated conversations through Facebook's application Messenger. The remainder were not engaged due to limitations in the number and capacity of intervention providers within this pilot programme, underscoring the need to train and professionalise intervention providers as well as explore technological innovation that can enable online outreach with greater ease and at greater scale.

Three metrics were used to consider the impact of online outreach: initial response rates, sustained engagements (conversations that included five or more messages between the candidate and intervention provider), and indications of potential positive impact during the course of the conversations.

Overall just under one in five (20%) candidates who were contacted responded, a significantly higher response rate than that usually seen with unsolicited email campaigns. Islamist candidates were more likely to respond,[6] with a response rate of one in four

(26%) compared with one in six (16%) for extreme right candidates.

Sustained engagement rates between intervention candidates and providers were achieved with nearly three out of four (71%) Islamist candidates and nearly two out of three (64%) extreme right candidates.

One in ten (10%) sustained conversations suggested the programme had a positive impact. This included candidates:

- Expressing an interest to take their conversation offline;
- Indicating that the conversations had challenged or changed their attitudes or beliefs;
- Suggesting that the conversations had a positive impact on their negative online behaviours.

We found that intervention providers were most successful in achieving sustained engagement when they responded immediately to a candidate who had replied to them, adopted a casual or meditative tone, and explicitly mentioned and discussed extremism.

Intervention Providers

Success rates of intervention providers varied. The most successful intervention provider achieved a response rate of 46%, with 83% of those interactions being sustained, and the longest conversation including more than 500 exchanges, demonstrating the best practice potential of providers.

The pilot was designed with a view to testing the viability of different types of intervention provider. The findings suggest that a variety of types of intervention provider can be successful:

- Professional counsellors were able to deliver more conversations than former extremists and survivors of extremist violence.
- Survivors of extremist violence were most likely to have a sustained engagement
- Former extremists delivered the fewest number of conversations over the course of the programme due to other professional responsibilities, but were the most likely to get an initial response.

Implications and Recommendations

ISD's Counter Conversations programme provides promising evidence that a solid proportion of individuals expressing support for violent extremism online can be identified at speed and scale and encouraged to engage with online intervention providers on a sustained basis.

There is potential to scale this form of online 'counter conversations' work across different social media platforms beyond Facebook, by expanding ISD's semi-automated identification technology and methodology, and scaling up intervention providers, including training professional counsellors so that they are confident to deliver this type of work, as well as increasing support for former extremists and survivors of extremist attacks through ISD's AVE Network.

There are however also significant risks inherent in online outreach work, particularly when targeting individuals already displaying signs of radicalisation. These must be taken into account in the design of any scaled, professional programme:

- **De-confliction:** A secure process of 'de-confliction' should be considered to avoid online outreach with individuals who are subjects of active police or security service investigations.
- **Automation:** While some form of automation is necessary to scale, it is vital that any automated identification process is supplemented with expert manual review in order to minimise the risk of outreach with 'false positives' (individuals accidentally or wrongly identified as being extremist).
- **Operating in countries that lack human-rights-compliant referral mechanisms:** Online intervention programmes should only operate in countries with human-rights-compliant referral mechanisms. If this is not observed then intervention providers who operate in these countries may face legal risk, and candidates face exposure to potential human rights abuses including unlawful detention, torture or even extrajudicial killings by police or security services.
- **Legal liability:** Online intervention providers must be mindful of local laws that apply to having contact with known extremists as this could expose intervention providers to prosecution.

- **Links with government-run programmes:** The public is likely to view government involvement in online outreach as problematic Moreover, in many countries governments' presence and ability to undertake this work is severely constrained by legislation like the Investigatory Powers Act 2016 in the UK.[7]

With appropriate risk mitigation – including clear ethical frameworks and safeguarding procedures – it is possible to create an effective system of direct messaging interventions at scale. To do so we make the following recommendations:

Leverage technology

Leverage technology to achieve scale in both identification and intervention across multiple social media platforms:

- Further refine and develop semi-automated identification solutions across alternative social media platforms.
- Explore additional applications of technology to enable a triage-like system for initiating conversations, based on innovative programmes such as Crisis Text Line, and the use of video chat services for in-depth face-to-face engagement.

Professionalise intervention

Professionalise intervention providers through: training, salary and pastoral support and use networks like ISD's AVE Network to provide a 'community of support' to formers and survivors interested in delivering interventions

- Develop an accredited training programme and qualification for intervention providers.
- Pay intervention providers and provide pastoral support; ISD's AVE Network provides an ideal framework for this.

Explore potential links

Explore potential links with NGO-operated offline 'exit', disengagement and intervention programmes:

- Explore how ISD's identification methodology can be applied to facilitate referral to offline intervention programmes.

Trial effectiveness

Trial the effectiveness of 'counter conversations' across the radicalisation spectrum:

- Trial supplementing counter-speech efforts with direct messaging counter-conversations further 'upstream' (i.e. to audiences who are not yet showing strong signs of radicalisation but who may be at risk, for example by being within friend networks of individuals promoting an extremist ideology).

Carry out further work

Carry out further work to evidence and identify what constitutes behavioural or attitudinal change following intervention to further define 'success':

- Explore how technology can be implemented to track instances of medium-to long-term behavioural change following online intervention.
- Explore the considerations surrounding applying offline and online interventions in countries that do not have human-rights-compliant referral mechanisms for intervention:
- Explore, within international institutions and organisations, the legal and policy frameworks for delivering counter conversations in high risk areas in an ethical and legal fashion.

> ISD's Counter Conversations programme provides promising evidence that a solid proportion of individuals expressing support for violent extremism online can be identified at speed and scale and encouraged to engage with online intervention providers on a sustained basis."

Background

This chapter provides background into the growth of de-radicalisation and disengagement initiatives offline over the past ten years, and outlines how ISD is working to bring this work into the online world.

Today, online communications channels and social media are increasingly seen as important realms where extremist ideas are created, shared, and take on a life of their own. The transnational and subcultural character of these groups on social media is also taken more seriously today than it was in the past. We take for granted that individuals in the online space can connect with like-minded people from all over the world, can form identities that are largely removed from their everyday life, and – most importantly – that these identities can in fact be more important than their offline one.[5]

On social media platforms like Facebook, Instagram, Twitter and Reddit individuals often connect with people miles away, learn to articulate and refine elements of their ideology, and could be pushed into violence.

The activities of extremist and terrorist groups on these platforms varies significantly, and includes the distribution of material which can help individuals prepare for terror attacks, the sharing of disinformation and propaganda material, the propagation of hate speech and cyber-bullying, and radicalisation and recruitment. These activities take place across open and closed fora, and although often prohibited by legislation they can also include grey-area material which gives rise for concern but does not break any laws. Whilst there have been concerted efforts by social media platforms to take down illegal content and content that contravenes platform Terms and Conditions, gaps remain for online counter-measures to challenge expressions of extremism when censorship is not possible or desirable.

While social media has altered the way in which recruitment to violent organisations takes place, there remain elements of this process which are fundamentally rooted in very human interactions between individuals.[6] In other words, while public posts and conversations are important for solidifying an individual's broader commitment to an organisation, real trust building, escalation of commitment and activity, and potential material support for terrorist groups tends to happen in more private and exclusive spaces.[7] In these private chats, real friendships are formed, trust bonds are established, and potential recruits obtain a deeper sense of importance and significance as they are now in direct contact with the leadership of organisations. The shift from public platforms to private chats, then, often reflects the inner shift that simultaneously occurs for individuals from simply supporting a group to becoming a real member of an organisation.[8]

With increasingly high profile attacks and plots by both the Islamic State and the extreme right in Europe and North America, policy makers must find ways to intervene with these individuals before they require the attention of law enforcers. While there have been many recent efforts to upscale production and dissemination of counter-speech and alternative narratives on social media, there have been very few attempts to initiate programmes of direct online engagement, which re-appropriate the tactics used by extremist recruiters with the aim of de-escalating individuals openly espousing violent extremism. ISD's Counter Conversations programme was designed to fill this void: to use direct, online outreach to see if radicalised individuals could be engaged in conversations with the aim of initiating a process of de-radicalisation and disengagement. This process is flexible and has the benefit of being able to engage individuals who may not have broken the law, or indeed the terms of service of a platform, but who nevertheless are showing signs that they are supporting ideologies and groups which can lead to violent and illegal behaviour.

The Rise and Mainstreaming of De-radicalisation

De-radicalisation and disengagement initiatives have become increasingly established offline over the past decade. They are now delivered with varying degrees of success in a number of countries including the UK, Denmark, France, Germany, the Netherlands, Saudi Arabia, Sweden and the USA. They are led by personnel in civil society organisations, local authorities and law enforcement and target a range of individuals, from those showing signs of radicalisation but who have not committed any crimes, all the way to individuals who have been convicted of terror offences.

Although models for de-radicalisation and disengagement differ – from group-led approaches to one-to-one engagements – the methods used are often influenced by the fields of public health and social work. Common practices include referral systems and telephone hotlines for concerned practitioners and members of the public; training for frontline practitioners to recognise indicators of potential risk; and multi-agency referral teams involving police, social workers and mental health support workers to review referral cases and recommend appropriate support.

Approaches to de-radicalisation and disengagement have been influenced by other intervention fields, including most notably programmes that aim to prevent youth crime or gang involvement, suicide, drug and alcohol misuse, and other risky behaviours.[9] This is partly because many of the vulnerability factors that underlie other social harms can also be seen in cases of potential radicalisation.

De-radicalisation and disengagement initiatives have adopted multi-layered approaches, based on core public health principles of primary, secondary and tertiary interventions. These include making holistic assessments of individual and social drivers of radicalisation, and attempting to address these negative influences through one-to-one, family-based and community-based interventions.

In order to draw on pre-existing infrastructure and support systems, de-radicalisation and prevention programmes are often included in programmes that deal with safeguarding around other social harms – particularly of young people. In Denmark, the schools, social services and police model has been in operation since the 1970s to reduce youth crime, and was adapted for prevention of radicalisation in the 2000s.

In the UK, mainstreaming of prevention has taken place through government legislation creating a Prevent Duty for statutory frontline workers, requiring them to have due regard to indicators of potential vulnerability to radicalisation, alongside their other safeguarding duties relating to abuse, female genital mutilation, drug and alcohol abuse, or mental health issues. Elsewhere, the inclusion of prevention work into existing social work support structures has been more pragmatic and locally driven, based on a belief that social workers often have the best skills and training to support individuals and families who are experiencing concerns related to radicalisation.

While de-radicalisation and disengagement intervention programmes have become more widespread in recent years, their true scale and impact is rarely if ever discussed publicly because of sensitivity and privacy concerns. This lack of transparency makes it difficult for researchers and practitioners to compare approaches and impact measures in order to understand what works. It also makes it difficult to compare the growing sector of interventions work in the context of de-radicalisation and disengagement with other more established intervention fields, including gang prevention, where there is a more robust and substantial academic literature on impact.

Practitioners and governments often attest to the success of de-radicalisation and disengagement interventions. The UK Home Office recently released statistics for its Channel programme. According to the government, over 300 individuals have successfully completed a Channel intervention with improved vulnerability, though precisely how this is measured – and what medium- to long-term monitors are in place to ensure that positive impact is sustained – remains hidden from the public realm. As a result of this lack of transparency, there is still controversy among some social scientists, public health practitioners and governments over the validity of de-radicalisation and disengagement approaches.

Types of intervention[10]

- **Primary**: Work with individuals at risk of radicalisation, with activities such as one-to one mentoring, psycho-social care and group interventions.
- **Secondary**: Work with immediate families and friends closely connected to an individual known to have engaged with violent extremism, with activities such as family counselling, welfare and general support.
- **Tertiary**: Activities with an individual's wider spheres of influence such as institutions and the wider community, which often seek to engage, protect or divert individuals who may be exposed to heightened risk factors.
- **Rehabilitation**: Working with high-risk groups (e.g. offenders while in prison, on remand, returnees and foreign fighters) to disengage, and where possible, de-radicalise them away from extremism.

> “Many of the vulnerability factors that underlie other social harms can also be seen in cases of potential radicalisation.”

Bringing Intervention Programmes Online

The opaque nature of current impact measures has made it difficult to establish expectations for what can be achieved through online interventions work. As this field is still very much in its infancy there is a dearth of literature on what constitutes successful practice. However, in recent years a number of social work and public health approaches have been used with social media and new technologies; examining them can add nuance to our understanding of this work.

Initiatives like Childline in the UK and Crisis Text Line in the US demonstrate the need for and potential of using new technologies and social media to support young people dealing with issues such as drugs and alcohol, gangs, bullying or suicide. These young people may feel more comfortable receiving anonymous text-based support from a trained counsellor rather than speaking to family and friends.[11]

In recent years, with the high profile nature of Isis-inspired attacks in the media, many of these programmes designed to support young people receive messages on extremism issues. However, the scale of extremism-related referrals to more general youth-based online intervention programmes – and the procedures in place for dealing with them – remains unknown and understudied.

One of the few areas in which intervention programmes have been taken into the online space has been in the field of cognitive behavioural therapy (CBT), a school of psychotherapy that helps individuals deal with difficult periods in their life, and attempts to change the way they interpret and respond to negative events and emotions. Moving CBT online arose from an acknowledgement among professionals that youth struggle with emotional issues, but have limited access to CBT. Ease of use and anonymity are two benefits of online therapy identified by parents of adolescents with mental health problems,[12] and studies have shown that online CBT for dealing with anxiety and depression has proved effective and practical.[13]

While online CBT is not directly comparable to online direct interventions with individuals espousing violent views on social media, it does show that there is potential for this kind of work online, with further professionalisation and methodological refinement, and policy makers are now debating whether online hate speech can be countered in similar ways.[14] Furthermore, the benefits of online CBT are transferable to extremism-related intervention: offline de-radicalisation work takes significant effort to bring individuals to the point that they are prepared to enter an intervention, but online outreach can be delivered quickly and with relative ease.[15]

Developing Direct Online Engagement to Counter Violent Extremism

In 2015 ISD conducted a small-scale pilot of one-to-one engagements, which demonstrated that direct, private and personalised online outreach is a viable mechanism to engage individuals who openly espouse support for violent extremism, and that these conversations may be useful in creating doubt in the minds of radicalised individuals.[16]

Furthermore, it became apparent that direct online engagement could be a valuable method for engaging hard-to-reach individuals who would otherwise be inaccessible to de-radicalisation infrastructure, and whose support for violent extremism may not be apparent through their offline behaviour.

However, these findings were based on a small sample size of approximately 150 individuals and provided little insight into how it might be possible to systematise and formalise a methodology for direct online outreach as a counter-extremism methodology.

In order to further understand the efficacy of direct online engagement we conducted this second phase of the One to One programme, the findings of which we present in this report. Here we intended to discover and establish what frameworks needed to be established in order to deliver outreach safely, responsibly and at scale. We also sought to deliver a large number of online engagements in order to understand what they are capable of achieving, and what factors can influence their outcomes.

It is important to set clear expectations for what online engagements can achieve at the outset. In the same way that an individual's journey to radicalisation is not the sole product of online influence, their exit and diversion will likely require more than just an online conversation. While possible, it is unlikely that an online conversation alone will lead to significant and measurable disengagement and de-radicalisation. What is more important at this early stage, we would argue, is simply to engage in a conversation that allows the topic of extremism to be broached and a positive alternative to extremist ideology to be proffered. Ideally these conversations can also progress in a manner that increases the chances of an individual agreeing consent to be directed to additional support structures such as mental health services, or a face-to-face meeting with a trained intervention provider. This simple engagement provides a unique window of opportunity to open trusted communications with individuals least likely to be known to frontline services or agencies and who are engaging with the most prolific extremist content online.

It also presents us with a critical and currently unexplored opportunity for credible intervention providers to challenge individuals on the views they are expressing safely.

In the sections below, we present the findings of our online outreach and engagement work. In the next section we present our identification methodology as well as some demographic data regarding the individuals identified as showing signs of radicalisation. Due to the sensitivity of this work, and in order to protect the confidentiality of both the candidates and the intervention providers, we are not able to reveal all of the details of the conversations that took place. We present the overall findings from our online interventions in terms of response rates and the efficacy of different intervention providers, before concluding with a series of recommendations for scaling this work to meet the scale of the challenge we face.

> “Engagement provides a unique window of opportunity to open trusted communications with individuals least likely to be known to frontline services or agencies and who are engaging with the most prolific extremist content online.”

Identification and Candidate Demographics

This section presents an outline of our semi-automated identification methodology, which resulted in the identification of over 800 individuals openly displaying signs of radicalisation, as well as the demographics of these candidates for engagement.

A core component of the One to One programme is the identification of individuals demonstrating support for extremist ideologies based on discernible, publicly available, online behaviour. This process needs to be robust and include safeguarding protocols to not only ensure privacy protections but also serve as a mechanism whereby individuals who are breaking the law or represent an immediate security risk can be referred to law enforcement. A full outline of safeguarding protocols and ethical guidelines for conducting this work are provided in the Appendix.

The rise of 'social listening' and online analytic tools has enabled researchers and counter-extremism practitioners to reveal the true scale of hate and violent speech online. This software can therefore be a valuable tool for identifying individuals displaying support, endorsement or encouragement of violent extremism online.

Yet, too often these tools have been developed with commercial purposes in mind, not for the fields of social science and public policy. A review of over 20 pieces of commercial social listening software revealed that none were flexible and open enough to be suited to the task of identifying individuals displaying signs of radicalisation online on Facebook. ISD thus developed a custom-made identification solution that combined NLP and machine learning in an open and transparent manner to enable continuous review by ISD radicalisation practitioners and experts.

A five step identification process was initiated, involving a combination of technological automation and manual verification by ISD's radicalisation experts and practitioners. It involved:

1. Identifying public Facebook pages that, although not necessarily violent in themselves were associated with Islamist and extreme right-wing groups or that tended to attract supporters of these ideologies, and identifying users who were posting comments on those pages. In this fashion over 42,000 individuals were identified, an overwhelming majority of which were extreme right;
2. Analysing comments posted on those public pages using a Natural Language Processing algorithm, searching for violent, aggressive and dehumanising language, and language associated with extremist ideologies. In this way the pool of individuals above was distilled to a group of around 7,000 individuals, of which we chose a sample of 1,600 for further examination;
3. Manually reviewing other indicators drawn from open source data to determine risk of radicalisation;
4. Assessing individuals against a risk matrix developed by ISD and based on a decade of in-house de-radicalisation experience, with risk ranked in four categories:

- **Intent:** the extent to which a candidate believes in and supports violent extremist ideology (e.g. if they comment on an extremist image in a positive fashion);
- **Engagement:** the extent to which an individual is engaged in extremist activity (e.g. if they regularly share extremist content);
- **Capability:** whether a candidate has the capability to commit a violently extreme act (e.g. if they post images of themselves with weapons, or indicate that they have a history of violent behaviour);
- **Support:** how much a candidate engages with individuals not involved with extremism, and how many of a candidate's friends and family were involved in extremism (e.g. if they only engage with extremist individuals on their Facebook wall);

5. Where necessary referring individuals to law enforcement, followed by candidate selection and secure transfer to Intervention Provider. Ultimately we selected 814 individuals who were assessed as displaying signs of radicalisation in their online behaviour as candidates for online engagement.

Overall our methodology proved particularly successful for identifying individuals who appeared openly supportive of extreme right ideologies or groups. The identification methodology was less effective when identifying individuals openly supporting Islamist extremist ideologies, which has generally faced greater pressure and attention from governments and social media companies.

Demographics of Candidates: Age, Gender and Location

One of the benefits of using Facebook as the platform for engagement is the publicly available demographic data regarding users' gender, age and location. Demographic data are useful not only from a research perspective in understanding who these individuals are, but also to facilitate effective conversations, for example, by matching candidates and intervention providers according to age or gender.

ISD researchers analysed the demographics of the majority of individuals who were engaging with pages associated with extremism and using violent, aggressive language. The data revealed that 77% of Islamist candidates appeared to be younger than 30, and 72% of extreme right individuals older than 45.

Interestingly, this trend was even more pronounced in women: four out of five (81%) extreme right female candidates were aged over 45 and the exact same proportion of Islamist female candidates were aged under 30 (Figure 1 + 2).

We also examined the locations of extreme right and Islamist candidates, which can be valuable when attempting eventually to link online interventions with either offline intervention providers or other statutory services within the area.

For example, in the UK it is notable that there are concentrations of both ideologies in key urban centres including Greater London, the West Midlands and the North West (Figure 4). However, as Figure 3 shows, extreme right candidates had a broader geographical distribution than Islamist candidates. This geographical information could potentially be useful to facilitate referral to other statutory or locally based services or programmes, which we recommend in the conclusions to this report. At the very least, when establishing better links between online outreach and offline intervention programmes, it can be valuable for statutory services to be aware of the information that can be delivered through adding an online component to their work.

Finally, in addition to their age, gender and location, users often publicly share further information regarding their interests or employment. Knowing an intervention candidate's wider interests – whether their favourite football team, sport or movie – can be a useful tool to build rapport and trust.

This data can demonstrate concerning trends that policy makers need to pay attention to. For example, our analysis of the profiles of extreme right males suggested that nearly two dozen – or 9% of the total of extreme right men – appeared to be current or former members of the armed forces.

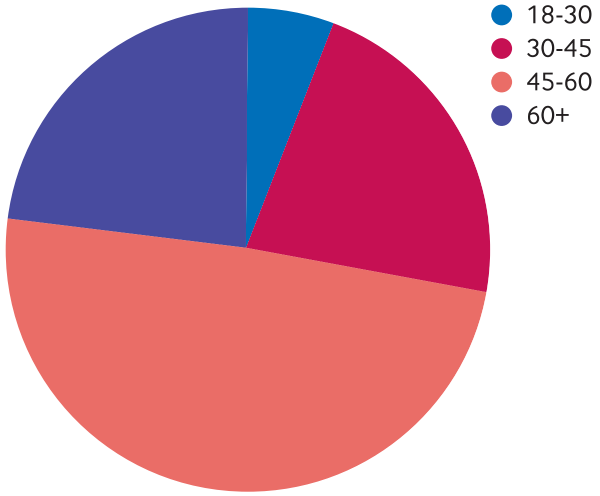


Figure 1.
Age breakdown of extreme right intervention candidates

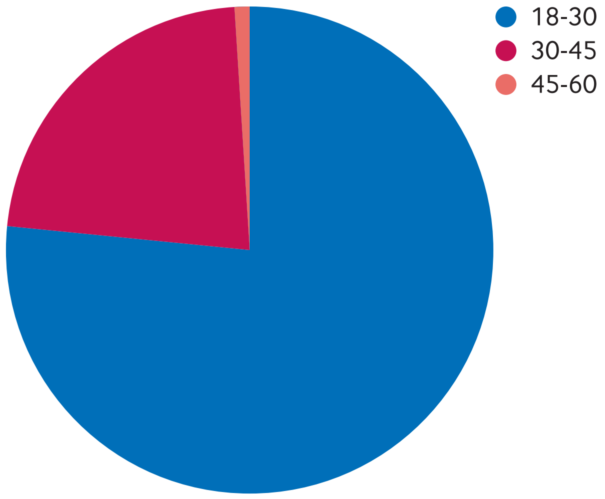


Figure 2.
Age breakdown of Islamist online intervention candidates

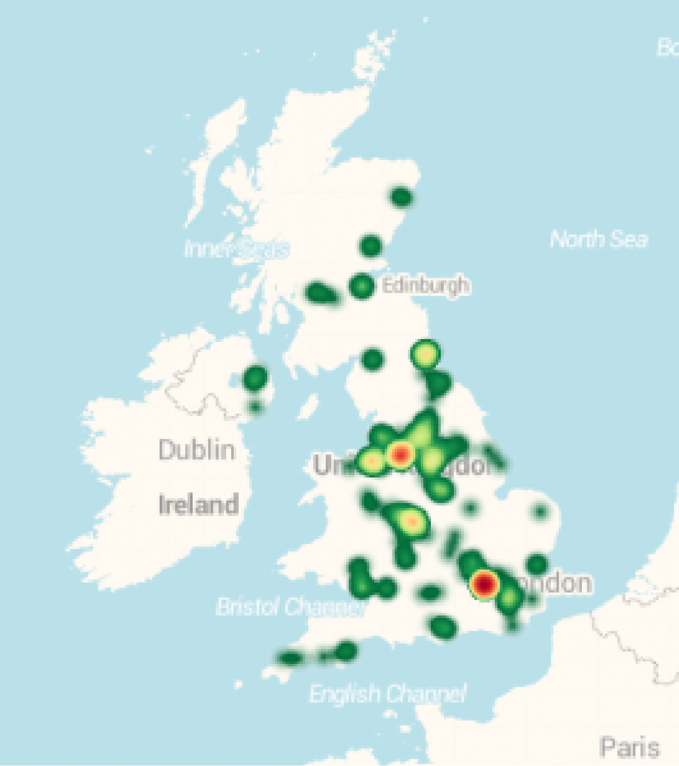


Figure 3.
The geographical distribution of extreme right online intervention candidates in the UK

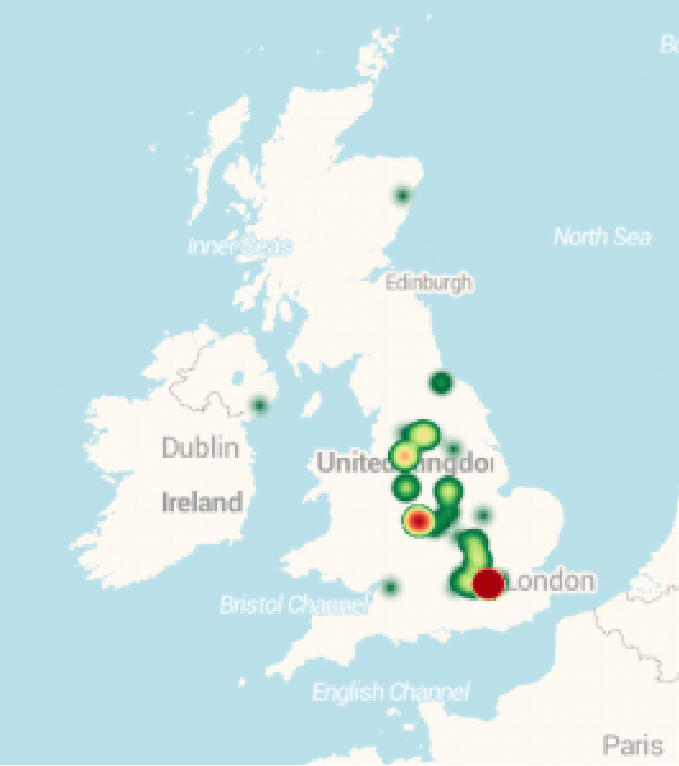


Figure 4.
The geographical distribution of Islamist online intervention candidates in the UK

Findings: Interventions and Engagement

This section outlines the lessons learned from our scaled programme of online outreach, including the three indicators used to measure success, the effectiveness of different types of intervention providers, and the impact of message tone, timing and content.

It is important at the outset to establish the right expectations for what online outreach can accomplish and what constitutes success. While it is possible that a conversation online could lead to de-radicalisation and disengagement, it is more realistic that this process would need to take place over multiple conversations, including either face-to-face in person or through interactive video technology. Moreover, an online interventions programme will always have to contend with whether someone's online behaviour is indeed an accurate reflection of their offline behaviour. As we argue in the recommendations section below, it may be possible to measure whether an individual continues to post extremist content on a platform like Facebook. However, the ability to engage with other social media platforms anonymously poses challenges to developing medium- and longer-term measures of impact.

While there is therefore scope for developing more sophisticated approaches to measuring the impact of online intervention work, at this stage of still demonstrating proof of concept, we used three metrics for determining success: initial response rates, indicators of sustained engagement and indicators of potential positive impact.

First, **initial response rates** – whether or not a candidate responded to an outreach attempt – allowed us to assess the validity of online outreach as a means of engaging at-risk individuals. The initial outreach undertaken by an intervention provider is most analogous to unsolicited emails and marketing campaigns, which according to MailChimp's email marketing benchmarks have an average of just 3 per cent engagement. Of course, without at least an initial reply, a system of online intervention is not possible.

The second measure we used was indicators of **sustained engagement**, which included conversations with five or more exchanges with a provider. Beyond generating an initial response, a programme of online interventions requires a willingness among intervention candidates to enter into a sustained conversation with an intervention provider. At first, the aim of these conversations should be to build trust between a provider and an intervention candidate, much in the same way that offline interventions require a period of trust building before progressing to either the content of extremist ideology, or the apparent drivers of an individual's radicalisation.

Third, having established the viability of facilitating sustained conversations, we sought to identify conversations that had qualitative indicators of **potential positive impact**, including suggestions that a conversation may have changed an individual's mind, admission that their online behaviour may be harmful to others, or requests to continue a conversation on another medium.

Overall we identified over 800 individuals who were displaying signs of radicalisation and initiated online interventions with just over half of these individuals. In total, conversations were initiated with 569 individuals, of which 112 responded to online outreach, 76 individuals engaged in a sustained conversation, and 8 of these sustained conversations showed indicators of potential positive impact. Although these numbers may seem small they nevertheless demonstrate that once individuals are engaged in a counter conversation this can be leveraged to generate positive impact in a significant proportion of cases. Due to the nature of online identification it is probable that a number of the individuals engaged in this programme would otherwise be inaccessible to pre-existing counter-radicalisation infrastructure, and without this work would be unlikely to encounter any support around these issues.

Response Rates

When considering response rates, it is important to emphasise that the initial messages from the intervention providers were unsolicited, private messages sent by individuals who were strangers to the intervention candidates. With that in mind, the fact that one in five intervention candidates replied to the initial message should be taken as a measure of success. It also underlines the importance of achieving scale, particularly at the identification stage.

Moreover, there was a notable difference in response rates between Islamist and extreme right candidates. Islamist candidates were more likely to respond to an intervention provider than their extreme right counterparts: one in four (26%) of Islamist candidates compared with one in six (16%) extreme right candidates did so. In both instances men were more likely to respond than women: 27% of Islamist men responded compared with 21% of women, and 21% of extreme right men responded compared with 9% of women.

This greater level of responsiveness among the Islamist candidates may be more a reflection of age than anything else. We found that older candidates were less likely to respond than younger candidates, potentially reflecting their technical literacy, which may explain why the extreme right, as the elder demographic engaged in this programme, were less likely to respond than the younger Islamists.

Sustained Engagements

Beyond an initial response, one of the key aims of the intervention providers was to initiate a conversation that was sustained, as a sustained conversation demonstrated a critical willingness among intervention candidates to engage. It also allowed the intervention providers to test out different methods of building trust with intervention candidates and drawing them into a conversation that could ultimately shift their attitudes and behaviours.

Again, there was an interesting difference between the ideologies, with Islamist candidates more likely than extreme right candidates to have sustained engagements: 71% of conversations with Islamist candidates were sustained compared with 64% among extreme right candidates. We also found that the ability to generate a sustained engagement seemed largely to relate to individual intervention providers. The most successful intervention provider had a sustained engagement rate of 83%. This provider's sustained engagements were also significantly longer than other intervention providers, with their longest conversation including more than 500 exchanges.

In comparison, several intervention providers who delivered a comparable or greater number of attempted conversations received sustained engagement rates ranging from 5% to 12%.

Indicators of Potential Positive Impact

In addition to measuring conversations based on length, we also qualitatively analysed conversations to detect instances which suggested that the conversation had possibly generated positive impact by presenting a candidate with an alternative point of view.

It is important to bear in mind that a candidate apologising for their comment or expressing regret could be disingenuous. Nevertheless we would still suggest that the internalisation of alternative arguments is an important process in diversifying an individual's perspective, a concept which is relevant to the 'contact hypothesis' framework in social psychology, which suggests that interpersonal contact with a diverse group of individuals is an important step to reducing prejudice and increasing empathy.[17] This conception has been reinforced anecdotally by former extremists drawn from our AVE Network, who overwhelmingly suggest that a seed of doubt sewn in the mind is hugely important in initiating processes of de-radicalisation and disengagement.

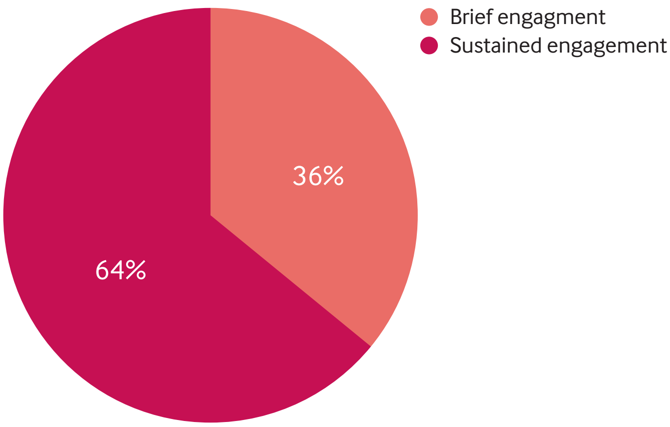


Figure 5.
The responses of extreme right candidates to outreach attempts

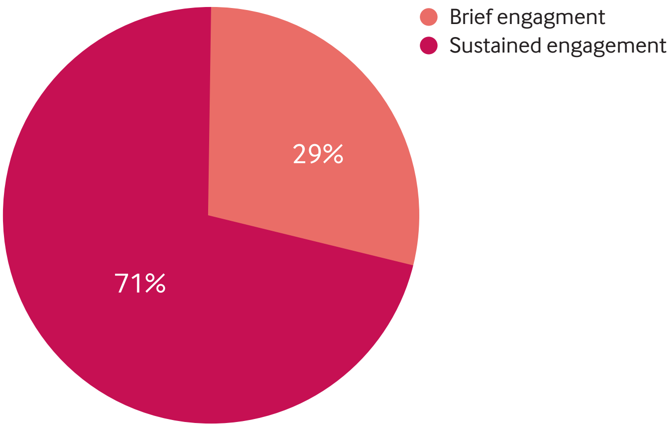


Figure 6.
The responses of Islamist candidates to outreach attempts

Additional monitoring and evaluation measures can help to elucidate the true impact of these indicators. In this programme we noted that two in five individuals who suggested that the conversation may have changed their mind ceased to post extremist content on their wall, but it is possible that the process of engagement may deter individuals from engaging in public fora in an extreme fashion while continuing to engage with extremist material in private. It is difficult to establish causality between observed moderation of shifts in behaviour and the delivery of an intervention, and we would suggest that more research needs to be conducted into measuring the impact of interventions.

We also counted a number of instances where candidates expressed a desire to move their conversation to another medium (such as a phone call or a face-to-face meeting). This is seen as an important measure as this suggests that online outreach can be a viable mechanism for bringing at-risk individuals to alternative areas of support such as offline counselling. We argue that this is the metric of possible success which has the potentiality to generate the most impact. If online conversations can be used to channel hard-to-reach individuals to offline support services there is the potential for significant prolonged engagement, where behavioural change can be evaluated through more established measures.

In total we found that one in ten sustained engagements explicitly contained indicators of potential positive impact. Similar to the findings outlined above, we found that conversations with Islamist candidates were more likely than those with extreme right candidates to suggest a potentially positive impact, with 7% of Islamist conversations and 6% of extreme right conversations yielding such results.

We also found that certain intervention providers were more likely to generate potential indicators of positive impact than others. An analysis of these more 'successful' intervention providers suggests that there are certain conversational approaches which result in longer engagements with richer discussion. It is apparent that a non-judgemental humanising approach to outreach results in individuals opening up about contentious issues. The intervention provider who consistently produced the longest and most impactful conversations would open a conversation by expressing interest in the extremist topics the candidate was publicising (e.g. saying 'Could you tell me more about what you are sharing …'), before presenting a non-judgemental opposing point (e.g. saying 'I just don't understand why there is so much violence in the world'), which could then be leveraged to generate a discussion. Intervention providers who were more forceful in their arguments also proved successful at generating responses, but these conversations were less likely to be coded as positive. In comparison, intervention providers who kept their conversations generic and vague (e.g. saying 'I can see you're passionate about Jane Austen') were much less likely to generate long or impactful conversations, suggesting that conversations should be kept on topic and discursive.

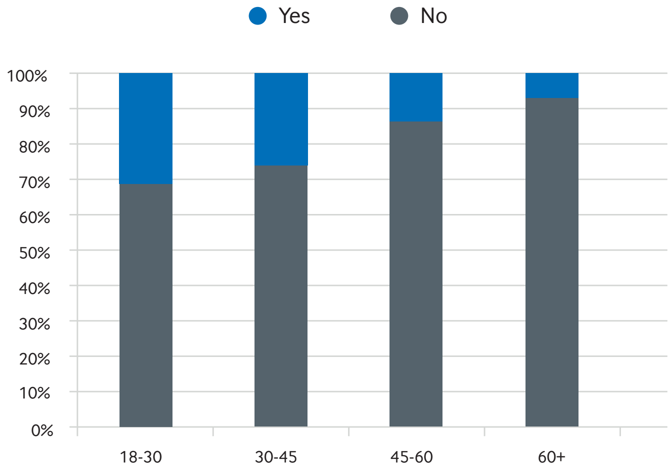


Figure 7.
Whether potential candidates responded to an initial attempt at outreach, by age

When considering the length and impact of conversations the demographics of candidates should be taken into account, as well as the stylistic approach of intervention providers. An analysis of candidate age reveals that individuals aged 18–30 are over four times more likely to respond to an attempt at outreach than candidates aged 60+, and over twice as likely to respond than individuals aged 45–60 (Figure 6). As extreme right candidates tended to be older than Islamist ones this trend can explain the discrepancies in responses between these two groups. A possible explanation for this trend can be the greater tech-savviness of younger people; as outreach was performed in a 'cold call' fashion, and the providers were not friends with the individuals they were reaching out to, the messages were delivered to a second inbox. We hypothesise that older Facebook users may be less aware of this secondary inbox and thus less likely to read an attempt at outreach.

Message Tone and Content

Conversations were qualitatively coded by tone and content in order to assess whether the construction of a message affected its likelihood of generating positive impact on the three categories cited above. Table 1 provides a summary of the different types of message tone and content used. Full descriptions are provided in the Appendix.

Message tone	• Argumentative • Meditative • Scholarly • Reflective • Sentimental • Casual
Message content	• Highlighting consequences of negative actions • Personal question • Ideological challenge • Personal story • Offer of assistance • General question • Mention of a shared interest • Mention of extremism • Mention of the programme

Table 1.
Types of message tone and content

Intervention providers were encouraged to experiment in their outreach approach and to adopt the tones which they felt most comfortable using, with the result that some tones were used more than others. Accordingly the numbers we are examining are not consistent, thus making it difficult to assess whether certain message tones were better at generating responses than others.

Figure 8 shows the total overall number of message tones and responses generated by initial outreach. These results suggest that most widely used message tones (casual and meditative) seemed to be the most successful at generating responses. While this could suggest that messages delivered in a tone that intervention providers are most comfortable with are more likely to generate responses, it could also suggest that adopting a casual tone at the outset of an intervention is essential to build initial trust.

With the above caveat in mind, the data suggest that certain tones seemed to resonate better among certain audiences. An argumentative tone seemed to generate more responses among extreme right audiences (Figure 9), while a meditative tone seemed to generate more responses among Islamist audiences (Figure 10). Reflective tones seem to be the least successful overall, while casual tones were relatively successful and also the most used across both sprints.

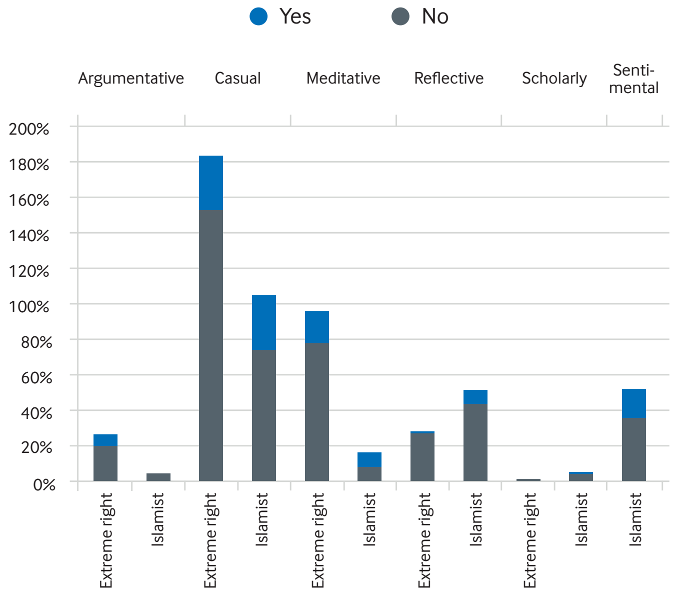


Figure 8.
Whether potential candidates responded to an initial attempt at outreach, by message tone

Moreover, there was no notable correlation between message tone and sustained engagement. These results suggest that the tone of a conversation seems to affect its impact, but the extent of this is limited.

When measuring the content of conversations we also found that certain topics were more likely than others to generate responses, particularly asking candidates personal questions, and highlighting the consequences of their negative actions, suggesting that tailoring conversations to be about an individual and their life is more likely to generate a response than conversations on more general themes.

Furthermore, conversations which explicitly mentioned extremism were more likely to generate sustained engagements than conversations that did not (Figure 12). Indeed, all of the conversations that had indicators of positive impact explicitly focused on extremist subject matter. However, the impact found from the explicit discussion of extremism was very different from the result when intervention candidates were made aware of the purpose of the intervention provider contacting them. Intervention providers were obliged to be transparent about their intentions and the project if asked by the intervention candidates; see the Appendix for further details. This occurred in seven instances, in each case leading the intervention candidate to either cut off communications or respond aggressively. These findings suggest that while individuals may react with hostility when they believe they are being subjected to an intervention, they nevertheless welcome the opportunity to enter into a conversation about extremist views to another person, which can then be potentially leveraged to open up a broader intervention.

This is an important factor as the delivery of counter-conversations which are not explicitly branded as interventions would be very difficult to deliver offline, demonstrating the value which is afforded by the relative lack of intimacy in online communications channels.

We also found that the timing of conversations was particularly important. Immediate response from an intervention provider significantly increased the likelihood of a sustained engagement, with delayed responses usually resulting in shorter conversations, or a loss of interest by the candidate.

The intervention provider who was quickest to respond to an initial message had sustained engagements over 80% of the time, and these engagements were significantly longer than those of other intervention providers, with the longest conversation consisting of more than 500 exchanges. Furthermore, this provider had the largest number of candidates expressing desire to take their engagement onto another platform. The feasibility for intervention providers to respond quickly to conversations is of course limited by their personal circumstances and other commitments. However the positive results which quick responses generated raise considerations around how to support providers delivering outreach on a more prolonged basis.

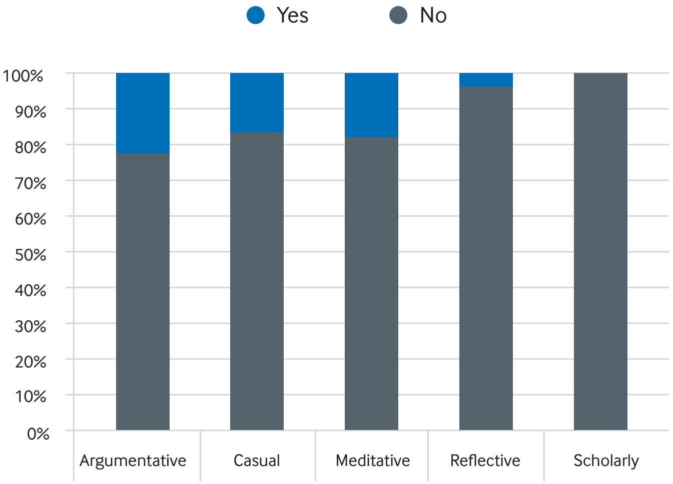


Figure 9.
Whether potential extreme right candidates responded to an initial attempt at outreach, by message tone

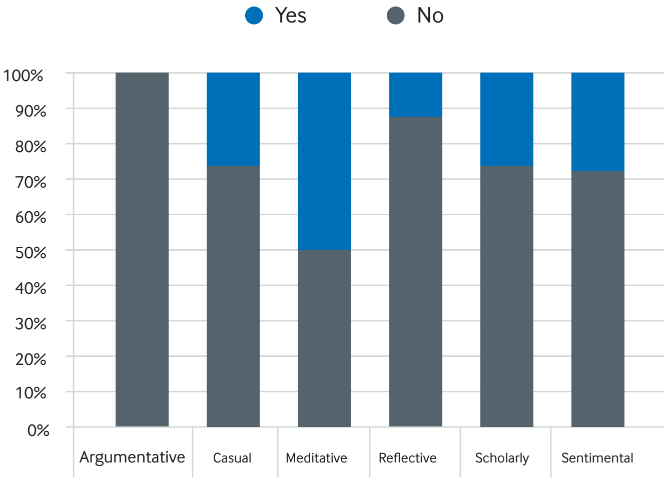


Figure 10.
Whether potential Islamist candidates responded to an initial attempt at outreach, by message tone

Intervention Providers

Over the course of the programme it was found that the ability to keep to outreach deadlines and deliver the target number of conversations varied according to each individual intervention provider (Figure 13). Overall we observed the following trends:

- Professional counsellors delivered the greatest number of conversations.
- Survivors of extremism generated the longest conversations.
- Former extremists got more responses to outreach requests than professional counsellors, but completed the fewest number of conversations.

These findings raise a number of considerations around the future implementation of this work. Due to other commitments, former extremists found the workload the most difficult to manage and accordingly delivered the fewest number of interventions. However their experience with the radicalisation process and with particular extremist groups proved to be valuable on a number of occasions. This would suggest that there should be a greater focus on building capacity amongst willing former extremists – and crucially a support infrastructure – to deliver this work.

Survivors proved to be very successful, and a number expressed their opinion of the value they felt in engaging with this work. However, as with formers, there needs to be a focus on building capacity amongst those who are prepared to work on counter-extremism initiatives. It is also critical that they receive ongoing and high quality pastoral care in conducting this work, which is very likely to emotionally draining.

Finally the ability of counsellors to manage their workload and engage at-risk individuals has significant implications for the scalability of this work. Although the personal experiences with and intimate knowledge of extremism brought by formers and survivors proved to be valuable resources for engaging at-risk individuals it should nevertheless be noted that the numbers of these individuals who are currently engaging or who are prepared to engage in counter-extremism work are limited globally, and not equally distributed across geographical and ideological contexts. However there are a great number of professional counsellors distributed globally. Accordingly this work suggests that through training counsellors it is possible to deliver interventions where there was previously no infrastructure to do so.

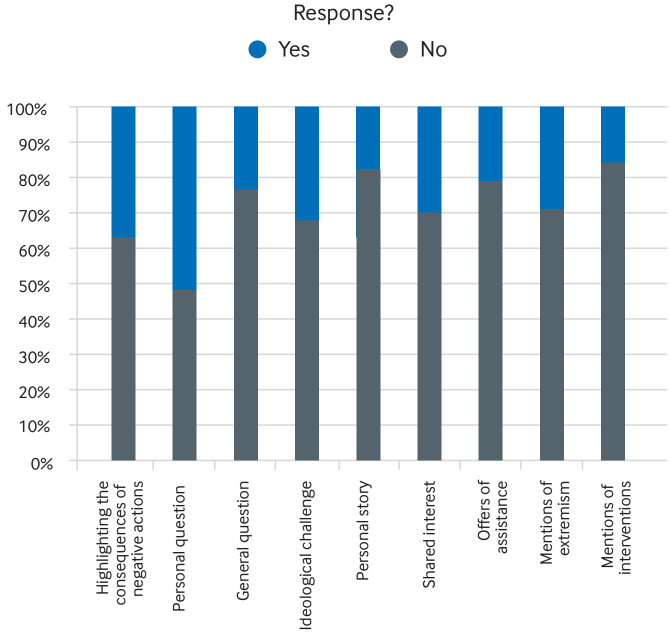


Figure 11.
Whether potential candidates responded to an initial attempt at outreach, by message content

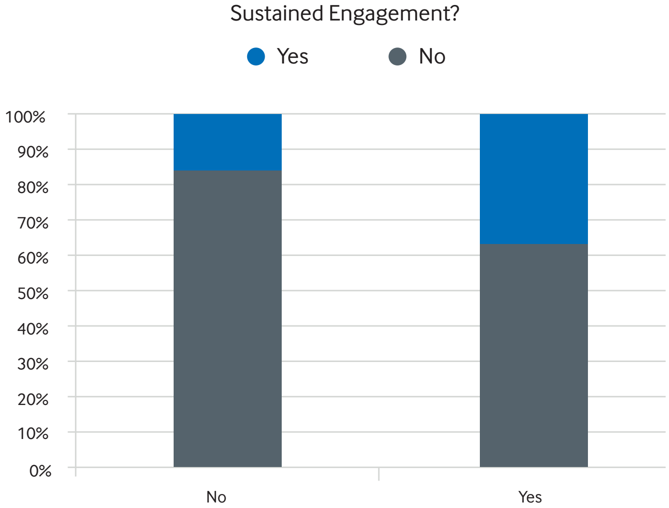


Figure 12.
The correlation between whether interventions mentioned extremism, and whether they were sustained engagements

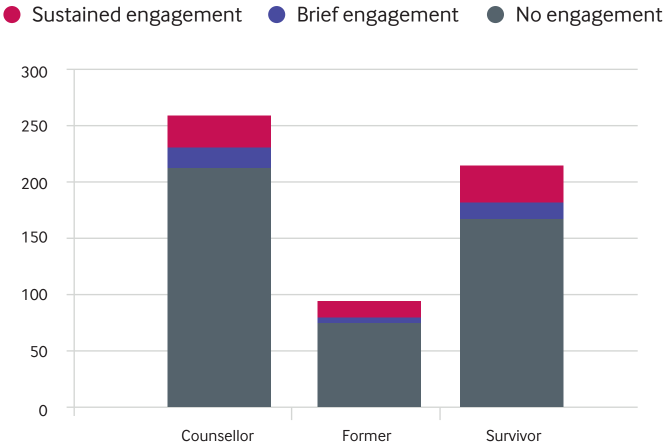


Figure 13.
Types of intervention providers and response rates

Conclusions and Implications for Future Programming

This final section outlines recommendations for policy makers on next steps for establishing a scaled, professionalised system for online interventions, including the risks that need to be considered and addressed.

The findings presented above underline the potential for professionalising and delivering one-to-one online direct messaging at scale. They show that individuals are prepared to enter into conversations online with strangers about their extremist behaviour, and that even very initial conversations can lead to potentially positive indicators of attitudinal shift, including a desire to continue an interaction face to face.

While the One to One programme has attempted to trial direct messaging 'downstream' with individuals showing signs of radicalisation, there is a need to develop the infrastructure to deliver direct messaging to individuals further 'upstream' – who are just beginning to show support for extremist narratives. While counter-speech initiatives are an important form of broadcasting to the many, undertaking direct one-to-one messaging – or 'counter conversations' – is vital to delivering the same personalised, peer-to-peer approach used by extremist groups.

Our online interventions already demonstrate a number of interesting and valuable insights that could be adopted by practitioners undertaking online interventions. For example:

- Islamist and extreme right groups operate differently online depending on their age. They also respond differently to outreach attempts.
- Outreach should be tailored to be about the individual who is being engaged and explicitly address their extremist behaviour. But a non-judgemental approach – which is core to social work – should guide the tone of intervention providers' conversations.
- Although the tone of the intervention provider did not particularly influence response rates, certain audiences seemed to respond better to certain tones. This should be explored in greater depth through deeper demographic analysis of different extremist ecosystems and further trialling direct messaging efforts.

However, before a system of online interventions can be established there are a number of risks that need to be addressed.

Risks and Mitigation

There are risks inherent in online outreach work, particularly when targeting individuals already displaying signs of radicalisation, which must be taken into account when designing a scaled, professional programme. These include:

- **De-confliction:** A secure process of 'de-confliction' is needed to avoid online outreach with individuals who are subjects of active police or security service investigations.
- **Automation:** While some form of automation is necessary to scale, a fully automated identification process could undermine online interventions if leading to outreach with 'false positives' (individuals accidentally or wrongly identified as being extremist).
- **Operating in countries that lack human-rights-compliant referral mechanisms:** Online intervention programmes should only operate in countries with human-rights-compliant referral mechanisms. If they operate in countries that have faced accusations of unlawful detention, torture or even extrajudicial killings by police or security services, intervention providers could be exposed to potential human rights abuses or significant legal risk.
- **Legal liability:** Online intervention providers must be mindful of local laws that apply to having contact with known extremists as these laws could expose them to prosecution.
- **Links with government-run programmes:** The public is likely to view government involvement in online outreach as highly problematic. Moreover, in many countries governments' presence and ability to undertake this work is severely constrained by legislation like the Investigatory Powers Act 2016 in the UK.

In addition to these risks, there is the possibility that online outreach could have the opposite effect than the one intended, by potentially hardening an individual's radicalisation or extremist mind set. While some of the individuals who we reached out to responded aggressively, or blocked the intervention providers, it is impossible to determine precisely what impact – if any – the outreach had on their views. As we recommend below, a scaled intervention system should aim to develop an approach to measuring impact that seeks to determine if each intervention candidate is either increasing or decreasing their online engagement with extremism (e.g. through further posting of violent, dehumanising language) following an attempt at online outreach. At the same time, simply allowing individuals openly to express support for extremist groups, and spread hateful, violent comments, is a risk in itself. Moreover, this same risk applies in offline intervention programmes, such as the UK's Channel Programme, and yet – as outlined in the background section to this report – these programmes now operate in a dozen countries and are seen as essential to preventing violent extremism.

Recommendations for Policy Makers

In order to be implemented effectively, a programme of counter-conversation interventions needs to be systematised and professionalised. To accomplish this, we make the following recommendations:

- **Leverage technology to achieve scale in both identification and intervention across multiple social media platforms.**

  The identification methodology deployed by ISD demonstrated the importance of combining technological automation with human analysis by subject matter experts. Simply automating online identification using commercial social listening software will lead to a large number of 'false positives', which risk undermining any programme of online direct messaging. Additionally, it is important to explore the utility of online interventions on alternative platforms which offer potential for direct messaging intervention at various stages, and which are known to be used by extremists, including Instagram, Twitter and Reddit. This is especially important given the migration of extremism supporters to alternative, encrypted platforms. At the same time, Facebook will likely continue to serve as a key platform for a scaled and professionalised interventions programme given the richness of its data and its large number of users.

  New technologies also hold the potential to facilitate the delivery of counter conversations at scale, for example through mobile phone applications. Organisations like Crisis Text Line demonstrate the potential for creating professional triage systems to assess risk and facilitate text-based interventions. Similar models can and should be applied to undertake direct messaging against hate speech and extremism, both 'upstream' and 'downstream'.

- **Professionalise intervention providers through training, salary and pastoral support and use networks like ISD's AVE Network to provide a 'community of support' to formers and survivors interested in delivering interventions.**

  Linking to community-led, grassroots organizations, such as the Against Violent Extremism (AVE) Network, the largest network of former extremists and survivors of extremist events in the world, would enhance intervention programming, by capacity building those who intimately understand pathways into, and out of, extremist and hate groups. The formers in AVE have been actively involved in interventions globally, as well as other educational initiatives, supporting counter and alternative narrative development, as well as law enforcement policy and practice. However, they lack the long-term support required to professionalize and scale-up their efforts, while providing them sustainable vocations post-disengagement.

  The findings of this programme demonstrate that further resources which equip intervention providers with specialist knowledge will likely increase their ability to deliver a range of interventions. Furthermore, the scoping of this programme has demonstrated the need to establish a number of support structures for intervention providers. The provision of psychological support to intervention providers was found to be particularly useful: this should be adopted as best practice by all organisations delivering this work. We found that intervention providers who are able to respond to a candidate immediately are much more likely to enter into a sustained and fruitful conversation with that person. This suggests that to deliver this work at scale it is necessary to establish a properly resourced, full-time unit of interventionists.

  It is vital to expand the pool of intervention providers to include social workers – as seen in both online and offline intervention programmes. However, training for social workers must include how to have difficult conversations on government policy, international geopolitics and religion and ideology. Similarly, formers and survivors need to have training in social work, trust-building techniques and talking therapy techniques, including on the importance of demonstrating a non-judgmental approach. Formers and survivors have an integral role to play in the intervention process, and their traumatic experiences within groups, provides them a deeper understanding of the trauma associated with being involved with extremist groups, as well as surviving, or losing loved ones, to extremist events. By providing them rigorous training, as well as subject matter expertise, formers would not only enhance their capacity to deliver interventions, but it would allow them to similarly have sustainable, and stable futures as therapists and counselors.

- **Explore potential links with NGO-operated offline intervention programmes.**

  Although online conversations show signs of potentially generating positive impact, it is difficult to maintain this unless online contact with an individual is prolonged for a significant period of time. While this does not necessarily have to take place in the 'real world' – for example, video chat applications could be used to engage face-to-face – possible integration with offline intervention programmes, particularly those that are NGO-led – should be explored. This could include using ISD's identification methodology to facilitate referral to offline support services. Further thinking needs to be done around the mechanisms and protocols that should be established to achieve this integration, especially where there is a strong government role in offline intervention programmes. Statutory and non-statutory bodies are often bound by different legislation around issues such as surveillance, for example, the Regulation of Investigatory Powers Act 2000 in the UK. If this work is to be integrated and scaled effectively it is imperative that a common mode of working is established between state and non-state actors that helps to ensure all the necessary compliance.

- **Develop and scale 'counter conversations' across the radicalisation spectrum.**

  This programme focused on delivering interventions to individuals who were demonstrating support for violent extremism on social media through their endorsement of extremist material, and their use of hateful and violent language towards others online. However there remains the possibility that the 'counter conversation' model can be employed across the radicalisation spectrum in order to support other counter-speech efforts. In particular it should be considered whether online outreach can be employed with audiences who are not yet showing strong signs of radicalisation but who may be at risk, for example by being within friend networks of individuals promoting an extremist ideology, or through expressing sympathy for extremist talking points.

- **Carry out further work to evidence and identify what constitutes behavioural or attitudinal change following intervention to further define 'success'.**

  Determining whether a change in online behaviour is mirrored in the offline world – without a formal link to an offline intervention programme or relevant statutory services – is impossible without contravening an individual's privacy. And it is important to note that offline intervention programmes often have to deal with a similar ambiguity when attempting to assess for sure if an individual has actually changed their mind. However, it may be possible to develop a comprehensive online framework for measuring longer-term impact at least relating to online behaviour. For example, the extent to which an intervention candidate continues to post comments on extremist pages could be measured periodically over time. However, if this type of framework is deployed or supported by national governments it needs to comply with key legislation, such as the Investigatory Powers Act 2016 in the UK, which outlines what constitutes surveillance online, and the EU's General Data Protection Regulation 2016.

- **Explore the considerations surrounding applying this methodology in countries that do not have human-rights-compliant referral mechanisms for intervention.**

  The delivery of this programme relied on the fact that there was an ethical, reliable law enforcement response in place to refer individuals who represented an immediate security risk. Should online interventions be delivered in a high-risk context – for example, in countries in the Middle East or North Africa – without such apparatus in place then it is likely that the delivery of the programme would be unethical. At the same time, online interventions have the potential to engage hard-to-reach individuals who are unlikely to engage with frontline services, so they may be a valuable tool in countries with less developed infrastructures. In 2018, ISD through the Strong Cities Network will produce a policy and practice model to deliver online interventions in countries with non-human-rights-compliant referral mechanisms.

Appendix: Methodology

Identifying individuals who are openly espousing support for violent extremist ideology on public social media platforms, and conducting outreach with them, is highly sensitive work that demands a careful, ethical approach that seeks to do no harm. ISD takes concerns around privacy, data protection, storage and security, and ethics extremely seriously and worked to ensure that best practice was followed throughout the design and implementation of the programme. The sections below outline in full the methodology and ethical guidelines that were used as part of the One to One programme.

Platform Choice

Facebook was chosen for the richness of the publicly available data on user demographics and user interests (here it should be noted that at no point in this study did ISD have any privileged access to Facebook data).[18] Compared with other platforms, Facebook offers greater potential to construct a more complete profile of a potential candidate prior to outreach, allowing an assessment to be built of their engagement with extremist-related groups, their support for extremist ideology, and any other information that might be useful to deliver outreach. This information helps to ensure that researchers and practitioners are confident that an individual meets the threshold for online intervention, reducing the potential for false positives. ISD collected data using Facebook's public application platform interface (API) which only allows researchers to collect data from public, openly accessible Facebook pages associated with a group or organisation. It is not possible to use software to collect data from individuals' personal profiles.

ISD researchers are now scoping methodologies for delivering online interventions on other popular platforms including Instagram, Reddit and Twitter, as well as on fringe social media platforms which are increasingly being used by extremist communities.

Identification Process

The identification methodology and process was designed to combine technological automation with human expert analysis to provide careful checks to ensure that individuals identified met the criteria for interventions. This methodology used software that combines Natural Language Processing and machine learning to enable researchers to code and analyse large quantities of data gathered from social media platforms.

Safeguarding protocols were designed and implemented to ensure there were appropriate privacy protections and risk mitigation procedures for what to do if they were observed to be breaking the law, or to represent an immediate security risk. The programme followed the ethical guidelines laid out by the British Psychological Society in 2013[19] and the recommendations of the Ethics Working Committee of the Association of Internet Researchers,[20] and ISD produced a comprehensive data handling and protection policy to cover this programme.

The identification stage took place in two sprints. The first focused on individuals expressing support ideologies. The second focused on individuals expressing support for Islamist ideologies. Both sprints aimed to identify an equal number of men and women.

There were five steps in the identification process:

- Step 1: Seed page analysis and initial identification
- Step 2: Data enrichment through NLP
- Step 3: Manual verification and construction of candidate risk profile
- Step 4: Practitioner-led risk assessment
- Step 5: Candidate selection and secure transfer to provider.

Step 1: Seed Pages and Initial Identification

The first step in the identification process was to identify public Facebook pages that may attract individuals displaying signs of potential radicalisation to violence. ISD researchers identified public Facebook pages that, although not necessarily violent in themselves, were associated with extremist ideologies or groups, or that tended to attract individuals sharing extremist content or using violent or aggressive language against others. Using Facebook's public API access, an initial data collection was conducted to identify individuals posting content on these pages initially identifying over 40,000 potentially at-risk users of pages associated with the far-right, and over 2,000 users of pages associated with Islamist extremism. This initial dataset was stored on a secure server, which had restricted password access only available to two ISD expert practitioners.

Step 2: Data Enrichment through Natural Language Processing

Focusing on the initial dataset of users gathered in Step 1, we then trained a NLP algorithm to examine user engagement with these pages, detecting instances of language that appeared to be violent, aggressive and dehumanising, or content keywords relevant to extremism, distilling the total pool of users identified in Step 1 to just over 7,000 individuals, with an overwhelming majority of these sitting in the far-right category. Due to programmatic capacity we selected a sample of 1,600 to be examined in further detail. We also developed a series of rankings and scores that used measures for the number of extremist-related pages the user posted on, the frequency of posts, and the extent to which their comments included violent or dehumanising language.

Step 3: Manual Verification and Construction of Candidate Risk Profile

In order to ensure the quality and accuracy of the data gathered in steps 1 and 2, ISD researchers conducted a series of manual checks on the users identified in Step 2. They assessed the language identified as violent or extremist in nature, and manually gathered additional information that demonstrated support for violent extremism from individuals' Facebook profiles. Relevant factors considered included:

- Profile pictures containing extremist imagery;
- Likes or positive comments on pieces of content supporting violently extreme groups;
- Likes or positive comments on extremist material;
- Posts supporting violently extreme groups;
- Posts containing extremist material;
- Friends within extremist networks;
- Indication of offline involvement with extremist groups.

This information was then collated and used to create a full profile of potential radicalisation risk. All information relating to individuals was stored securely on an encrypted drive to which only the four ISD expert staff working on the project had access. No hard copies of data were made over the course of this programme, and by time of publication all data which can be used to identify an individual will be deleted.

Step 4: Practitioner-led Risk Assessment

ISD's in-house de-radicalisation practitioners assessed individuals against a purpose-built risk matrix, and where necessary made referrals to law enforcement. This risk matrix was developed in-house, and was based on two pre-established government frameworks for mapping behavioural risk relating to extremism, which were adapted to examine online behaviour only: the Channel Vulnerability assessment framework and the Local Safeguarding Children Board London Threshold Document.[21] This step in the process helped to ensure that candidates selected for outreach met the threshold for intervention, and was critical to judging whether candidates presented an imminent risk of violence or criminality . The risk assessment process was performed regularly throughout the programme in order to adjust for any changes in candidate behaviour following identification, and to incorporate any additional information which became apparent during engagement. Behavioural risk was ranked in four categories:

- **Intent:** the extent to which a candidate believes in and supports violent extremist ideology (e.g. if they comment on an extremist image in a positive fashion);
- **Engagement:** the extent to which an individual is engaged in extremist activity (e.g. if they regularly share extremist content);
- **Capability:** whether a candidate has the capability to commit extremist violently act (e.g. if they post images of themselves with weapons or appeared to have a history of violence);
- **Support:** how much a candidate engages with individuals not involved with extremism, and how many of a candidate's friends and family were involved in extremism (e.g. if an individual only engages with extremist individuals on their Facebook wall it suggests the individual does not have support from anyone other than their Facebook contacts).

Each of these risks was ranked from 1 to 4, with 1 being the lowest concern and 4 being the greatest concern. Possible mitigating factors and possible enhancements to the risk factors were also noted, based on the threshold of needs matrix used in UK social work. A shortened version of this risk matrix is shown in Table 2.[22]

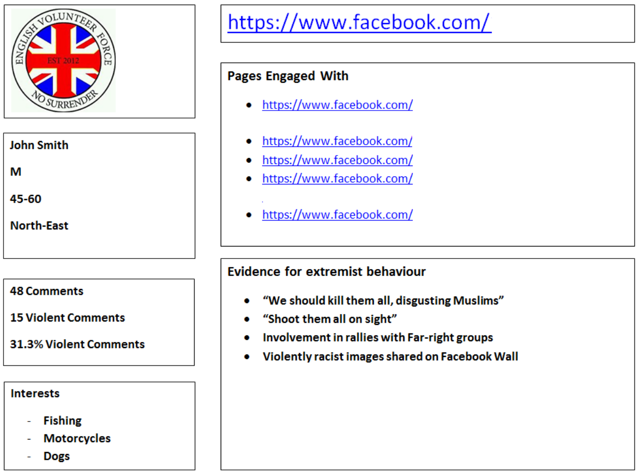
There was also a series of urgent indicators, which suggested whether a candidate posed an immediate risk to themselves or others.

Tier 1	Tier 2	Tier 3	Tier 4	Urgent criteria
No additional needs	**Early help**	**Multiple needs**	**Acute needs**	**Immediate Risk**
These individuals are not showing signs of radicalisation, and their needs are met by universal services.	These individuals have additional needs and may be vulnerable and/or showing early signs of radicalisation. Their needs are not clear, not known or not being met. This is the threshold for a multi-agency early help assessment to begin. These individuals require additional support, which can be provided within universal and targeted services.	These individuals are showing signs of radicalisation and require specialist services to achieve or maintain a satisfactory level of resilience. They may require longer-term intervention from specialist services.	These individuals are significantly radicalised and are suffering or are likely to suffer significant harm. For young candidates this is the threshold for child protection. This would also include those remanded into custody and statutory youth offending services.	The candidate expresses an immediate intent to harm themselves or others. The candidate has a direct and explicit connection to a proscribed terrorist group. The candidate is seen to be preparing or readying weapons that could be ued to harm themselves or others. The candidate is discussing leaving their family for a violently extreme or terrorist group. The candidate's family or close friends have a direct and explicit connection to an individual or group who has harmed, or is known to want to harm, others.

Table 2.
Shortened version of the Local Safeguarding Children Board London Threshold Document
Source: http://www.londoncp.co.uk/files/revised_guidance_thresholds.pdf

Step 5: Candidate Selection and Secure Transfer to Intervention Provider

On completion of Step 4, a risk profile was created for each intervention candidate selected for outreach and then encrypted to protect against any potential data breaches before being transferred to the intervention providers. The information provided included a hyperlink to the candidate's Facebook profile, a list of pages which the candidate had engaged with, any additional evidence of extremist behaviour, figures on the number of engagements he or she had had with extremist pages, and useful demographic information for use by intervention providers to tailor their engagement style (Figure 14).


ENGLISH VOLUNTEER FORCE
EST 2012
NO SURRENDER

John Smith

M

45-60

North-East

48 Comments

15 Violent Comments

31.3% Violent Comments

Interests

- Fishing
- Motorcycles
- Dogs

https://www.facebook.com/

Pages Engaged With

- https://www.facebook.com/
- https://www.facebook.com/
- https://www.facebook.com/
- https://www.facebook.com/
- https://www.facebook.com/

Evidence for extremist behaviour

- "We should kill them all, disgusting Muslims"
- "Shoot them all on sight"
- Involvement in rallies with Far-right groups
- Violently racist images shared on Facebook Wall

Figure 14.
A fictional example of an online candidate's risk profile

Identification Methodology Limitations

A key limitation of the identification methodology is its reliance on user interaction with 'seed pages' – public and open Facebook pages. Thus, if a certain ideology in a specific geography has a large number of Facebook pages associated with it, and these Facebook pages have a particularly active user base, then identification of candidates is efficacious.

Our methodology proved particularly successful to identify UK extreme right candidates. Reasons for this could include:

- There are a large amount of pages associated with extreme right ideology.
- These pages are often highly localised, increasing the likelihood that candidates are from the UK.
- Those using these pages are more engaged and more likely to interact with these pages than individuals visiting Islamist pages.
- Those using these pages do not fear repercussions and so are more likely to interact with them in an extremist fashion.

However, we should be aware that our sources for data may have skewed our results. As this methodology relies on data gathered from public Facebook pages, our source group contained a number of street protest groups associated with Far Right activity and violence, but a small number of openly white supremacist groups and alt-right groups. While this is relatively representative of the extreme right ecosystem in the UK, it is likely that the resultant group of candidates did not necessarily represent the full spectrum of extreme right ideology.

The identification methodology was less effective to distinguish Islamist than extreme right candidates. A major reason for this was that a majority of individuals engaging with the Islamist seed pages did not reside in the UK, and there are also some key differences in the use of Facebook by Islamist and extreme right extremists:

- There are fewer Facebook pages associated with Islamist than extreme right extremism.
- Islamist pages are less localised in their focus than extreme right pages, and more likely to attract a global user base.
- Individuals are less likely to engage with Islamist pages than extreme right pages in an extremist fashion, potentially due to fear of scrutiny and repercussion.

As a result of these discrepancies more traditional manual research was needed to identify Islamist candidates, using a 'snowball' technique to examine their behaviour online.

Training

The 2015 One to One pilot had no formalised training component. As the intervention providers identified in this programme all had varying expertise in countering violent extremism, and in order to make steps towards creating a replicable and scalable programme, we devised a training seminar covering the following modules:

- A short review of the One to One pilot project, exploring previous results;
- Background information on extremism and extremist content;
- The operational protocol and procedures for the programme and the technology used for the project;
- Risk assessment protocol and measurement and evaluation procedures;
- The safety and security protocol for the programme.

Post-training surveys taken at the close of the training showed that all ten intervention providers found the training provided them a 'good' or 'excellent' understanding of the programme, including the technology used to identify candidates, radicalisation dynamics and the operational processes of the programme.

Support Infrastructure

In addition to training, it was essential to provide psychological and security support to intervention providers.

Psychological Support

There is the risk that engaging with individuals who express support for extremist ideology, either through comments or sharing of material, may trigger an adverse psychological reaction in an intervention provider, particularly one who has had previous traumatic experiences as a result of extremism. Therefore regular supervision sessions were a compulsory component for intervention providers throughout the programme. They typically included counselling or psychotherapy sessions for those currently working in counselling or psychotherapy, usually carried out by an independent third party. This is established best practice for most professional counselling bodies within the UK, including the British Association for Counselling and Psychotherapy. Supervision is seen as an ethical imperative for organisations delivering counselling work, and helps secure the wellbeing of counsellors and add an extra check or balance in case of any serious oversight.

The provision of compulsory supervision was received positively by our intervention providers, and provided them with the opportunity to reflect on their work in the programme, and address the strains that regular engagement with extremist individuals and content caused them.

Security Support

Before the start of the project a security consultant was employed to build best practice around the security of intervention providers. This was conveyed in the training modules and focused on safe operational approaches including:

- Safe use of technology (including how to turn geo-location off on devices);
- Safe operational practice (including the use of a secure device);
- Safe communicating practice (including setting clear boundaries).

To ensure these measures were being met, and to provide an extra layer of safeguarding, ISD staff assessed all provider accounts and conversations daily.

Outreach Protocol

Once candidates for outreach were identified they were transferred to intervention providers who began conversations with them using the Facebook Messenger application. Figure 12 outlines the operational process for the whole programme.

Intervention providers were instructed to attempt outreach twice, and if they heard no reply to cease attempts to engage.

Measuring Conversations

This section outlines the methodology adopted to measure the impact of these engagements.

Response Rates

The first level at which an engagement was measured was whether or not a candidate responded to an outreach attempt. Responses to initial outreach provide an entry with which an intervention provider can engage a candidate about their behaviour, and validate direct outreach as a methodology.

Sustained Engagement

A sustained engagement is measured as five or more exchanges between an intervention provider and intervention candidate. Although sustained engagement does not give a direct indication of the content and tone of a conversation, it demonstrates that individuals who are expressing support for violent extremism online are prepared to enter into a dialogue about their views. Though this dialogue is not necessarily always constructive, particularly at the outset, it nevertheless shows that online outreach works as a way of accessing, engaging and offering support to radicalised and vulnerable individuals who may otherwise be inaccessible to frontline support services. As with intervention grammes of support, including de-radicalisation initiatives.

Coding Conversations

In addition to measuring conversations we also sought to qualitatively code them by tone and content in order to assess whether the way a message was constructed affected its likelihood of generating positive impact. All conversations were manually coded by two ISD researchers, with a third ISD researcher providing a blind coding review and check.

Message Tone

To ascertain whether certain approaches are more likely to be successful at having an impact, the message tone was coded into six categories:

- **Argumentative:** adopts a deliberately provocative approach by directly challenging the individual with the aim of effecting a response;
- **Meditative:** encourages consideration of a candidate's perceived views, or questions if and why a candidate may hold the views reflected through their online behaviour;
- **Scholarly:** makes a factual approach, usually formatted as a statement rather than question, referring to academia, history, theology, etc.;
- **Reflective:** aims to stimulate self-reflection and consideration of the impact a candidate's views and online behaviour may have on themselves or others around them;
- **Sentimental:** expresses concern for the candidate, and incorporates terminology that indicates an emotional response to the candidate's profile and content therein;
- **Casual:** introduces a general question or comment, often around a candidate's personal interests.

Message Content

In order to ascertain whether certain subjects are likely to be successful at generating impact message content was coded into nine categories:

- **Highlighting consequences of negative actions:** encouraging reconsideration of a candidate's online behaviour by stressing the impact it may have on him or herself, and on those in the candidate's network; the consequences considered must be specific to the candidate's current behaviour and/or perceived mentality, and not relate to negative actions of extremist organisations as a whole;
- **Personal questions:** any inquiry into a candidate's involvement in, experience with, or exposure to extremist groups, ideologies and narratives; those on a candidate's passions or their sentiments towards extremist groups do not qualify as personal questions;
- **Ideological challenges:** encouraging reconsideration of a candidate's alleged views through highlighting the negative consequences of the actions of groups who employ similar rhetoric, through emphasising inconsistencies that often exist in extremist narratives, and/or stressing the impracticalities and irrationalities that accompany hateful discourse;
- **Personal stories:** any story or personal detail from the intervention provider's past, usually relating to personal experiences with radicalisation, or the results of extremism;
- **Offers of assistance:** in many forms, from offering a chance to talk, to help exiting an extremist movement;
- **General questions:** about a candidate's passions or their attitudes to an extremist group not covered by the 'personal question' category;
- **Mention of a shared interests:** touching on common ground with a candidate by mentioning a shared interest;
- **Mention of extremism:** any mention of extremism or extremist ideology by the intervention provider;
- **Mention of the programme:** any conversation where the One to One programme was explicitly mentioned.

Intervention providers were encouraged to experiment in their outreach approach and to adopt the tones which they felt most comfortable using, with the result that some were used more than others. Accordingly the numbers we are examining are not consistent, so it is difficult to assess whether certain messaging is better at generating responses than others.

Transparency

In order to preserve programme transparency, intervention providers were requested to inform candidates that they were the subject of a counter-extremism engagement if asked why they were reaching out to them. Seven candidates asked this question and were informed that they were taking part in an outreach programme, with the following results: two blocked their intervention provider without further engagement, one responded aggressively before blocking their intervention provider, two stopped responding to the conversation, one briefly asked about the programme then stopped engaging in the conversation, and one continued engaging with their intervention provider but aggressively.

Although explicit mention of the conversation being an 'intervention' generated a negative response, conversations that explicitly mentioned extremism were much more likely to generate sustained engagements than those that did not. Furthermore, all conversations that had indicators of positive impact explicitly focused on extremist subject matter. This suggests that while individuals may be hostile to the programme's intentions, they nevertheless welcome the opportunity to talk about extremist views to another person. This lesson must guide the design of future online outreach programmes.

More advice needed?

The delivery of online interventions is potentially high risk and requires a balanced, ethical and secure approach to ensure intervention providers and candidates themselves are protected. If you are considering how to deliver an online interventions programme and would like to receive specific advice or help setting up such an initiative, please contact info@isdglobal.org.

Endnotes

1. *The AVE is the largest membership-based group of former violent extremists and survivors of violent extremist events in the world. Founded in 2011, by a team of 'formers,' survivors, the Institute for Strategic Dialogue, and supported by Google Ideas, the Gen Next Foundation, the network has been actively working to disrupt and degrade a range of extremist narratives, while providing avenues for at-risk youth and current extremists to disengage. Driving the engine of the network is the power of 'formers' and survivors and their ability to utilize their intimate knowledge of the push and pull factors that drive others to join a range of extremist organizations, as well as their personal journeys of struggle, pain, hate, and ultimately, transformation. The AVE Network has been innovating preventative and intervention programming globally in a range of contexts for the past six years. For more information please visit http://www.againstviolentextremism.org/.*

2. *Here it should be noted that the identification methodology relied on publically available data and that ISD received no preferential access from Facebook.*

3. *NLP is a field of computer science which focuses on the interaction between computers and human language. NLP algorithms allow for the analysis of large amounts of text. Machine Learning is a field of computer science which uses artificial intelligence to enable computers to 'learn' without being explicitly programmed.*

4. *A sample was taken due to capacity limitations on the number of individuals that could be engaged during the course of this pilot programme*

5. *It should be noted that age was calculated from information which individuals self-reported, and might not necessarily be accurate.*

6. *These response rates are significantly higher than those for unsolicited email campaigns, which average at 3.19%, and slightly lower than those recorded for cold-calling campaigns (28%): Mailchimp. Email Marketing Benchmarks (2017), https://mailchimp.com/resources/research/email-marketing-benchmarks/#section_average_by_industry; Lampertz, D. Has Cold Calling Gone Cold? (2012), https://www.baylor.edu/content/services/document.php/183060.pdf.*

7. *For more information on the Investigatory Powers Act 2016 please see: http://www.legislation.gov.uk/ukpga/2016/25/contents/enacted.*

8. *Alarid, Maeghin. "Recruitment and Radicalization: The Role of Social Media and New Technology". In Michelle Hughes and Michael Miklaucic (eds), Impunity: Countering Illicit Power in War and Transition (Center for Complex Operations, 2016), http://cco.ndu.edu/News/Article/780274/chapter-13-recruitment-and-radicalization-the-role-of-social-media-and-new-tech/.*

9. *Decker, Scott, and David Pyrooz. "I'm Down for a Jihad: How 100 Years of Gang Research Can Inform the Study of Terrorism, Radicalization and Extremism", Perspectives on Terrorism 9, no. 1 (2015), http://www.terrorismanalysts.com/pt/index.php/pot/article/view/405/html.*

10. *Weine, Stevan, and Eisenman, David. "How Public Health Can Improve Initiatives to Counter Violent Extremism" (National Consortium for the Study of Terrorism and Responses to Terrorism, 2016), http://www.start.umd.edu/news/how-public-health-can-improve-initiatives-counter-violent-extremism.*

11. *For more information on ChildLine and Crisis Text Line please visit https://www.childline.org.uk/ and https:/ www.crisistextline.org/ ;*

12. *Eichman, Debra. "Online Cognitive Behavioral Therapy for the Prevention and Treatment of Depression and Anxiety in Children and Adolescents", Master of Science thesis (Winona State University: 2012), https://www.winona.edu/counseloreducation/Images/OnlineCBTforPrevention.pdf.*

13. *Andrews, Gavin, Cuijpers, Pim, Craske, Michelle G., McEvoy, Peter and Titov, Nickolai. "Computer Therapy for the Anxiety and Depressive Disorders Is Effective, Acceptable and Practical Health Care: A Meta-Analysis", PLoS ONE 5, issue 10 (2010), http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013196; Cavanagh, K., Shapiro, D. A., Van, D. B., Swain, S., Barkham, M. and Proudfoot, J. "The Acceptability of Computer-Aided Cognitive Behavioural Therapy: A Pragmatic Study", Cognitive Behaviour Therapy 38, Issue 4 (2009).*

14. *UNESCO. Countering Online Hate Speech, UNESCO Series on Internet Freedom (2015), http://unesdoc.unesco.org/images/0023/002332/233231e.pdf; Strachan, Anna Louise. "Interventions to Counter Hate Speech" (GSDRC Applied Knowledge Services, 2014), http://www.gsdrc.org/docs/open/hdq1116.pdf .*

15. *Amarasingam. "What Twitter Really Means for IslamicState Supporters".*

16. *Frenett, R. and Dow, M. "One to One Online Interventions – A pilot CVE Methodology" (Institute for Strategic Dialogue: 2015), http://www.isdglobal.org/wp-content/uploads/2016/04/One2One_Web_v9.pdf.*

17. *Rose-Stockewell, Tobias. "How We Broke Democracy" (Medium, November 11, 2016), https://medium.com/@tobiasrose/empathy-to-democracy-b7f04ab57eee; Carr, Caleb T. "Social Media and Intergroup Communication", Communication (August 2017), http://communication.oxfordre.com/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-460.*

18. *This applies, of course, only to users who have not activated privacy settings.*

19. *British Psychological Society. Ethics Guidelines for Internet Mediated Research (2013), http://www.bpsorg.uk/system/files/Public%20files/inf206-guidelinesfor-internet-mediated-research.pdf..*

20. *Association of Internet Researchers. Ethical Decision Making and Internet Research: Recommendations from the AoIR Ethics Working Committee Version 2.0 (2012), http://aoir.org/reports/ethics2.pdf.*

21. *UK Home Office. "Channel: Vulnerability Assessment Framework" (2012), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/118187/vul-assessment.pdf; London Child Protection Procedures. Threshold Document: Continuum of Help and Support [2017], http://www.londoncp.co.uk/files/revised_guidance_thresholds.pdf.*

22. *Ministry of Children and Youth Services. The Vibe is Respect (2016), http://www.ctys.org/wp-contentuploads/2016/12/The-Vibe-Report-English.pdf.*
See also: Government of Albert. Alberta Gang Prevention Strategy (2010), http://www.assembly.ab.ca/lao/libraryegovdocs/2010/aljag/9780778586173.pdf.

ISD

Powering new
generations
against extremism

PO Box 7814, London
United Kingdom, W1C 1YZ
www.isdglobal.org