

# counter-speech examining content that challenges extremism online

Jamie Bartlett

Alex Krasodonski-Jones

October 2015

Open Access. Some rights reserved.

As the publisher of this work, Demos wants to encourage the circulation of our work as widely as possible while retaining the copyright. We therefore have an open access policy which enables anyone to access our content online without charge. Anyone can download, save, perform or distribute this work in any format, including translation, without written permission. This is subject to the terms of the Demos licence found at the back of this publication. Its main conditions are:

- Demos and the author(s) are credited
- This summary and the address [www.demos.co.uk](http://www.demos.co.uk) are displayed
- The text is not altered and is used in full
- The work is not resold
- A copy of the work or link to its use online is sent to Demos.

You are welcome to ask for permission to use this work for purposes other than those covered by the licence. Demos gratefully acknowledges the work of Creative Commons in inspiring our approach to copyright. To find out more go to [www.creativecommons.org](http://www.creativecommons.org)



Published by Demos October 2015

© Demos. Some rights reserved.

Unit 1, Lloyds Wharf  
2-3 Mill Street  
London  
SE1 2BD

[hello@demos.co.uk](mailto:hello@demos.co.uk)

[www.demos.co.uk](http://www.demos.co.uk)

This paper was funded by Facebook. The views expressed are those of the authors and do not necessarily reflect those of Facebook.

## CONTENTS

Acknowledgments	4
Introduction	5
Method	7
Recommendations	14
Ways forward	15
Notes	18

## **ACKNOWLEDGMENTS**

We would like to thank Facebook for their support, in particular, Rosa Birch, Parisa Sabeti Zagat, Ciara Lyden and Winter Mason. From Demos we would like to thank Loraine Bussard, Alessia Tranchese, Raphael Hogarth, Sophie Gaston and Daniela Puska. All errors and omissions are ours.

## INTRODUCTION

Facebook serves almost 1.5 billion people globally. Although the majority of people use the site for positive purposes, there are some who use the platform in negative ways. With that in mind, Facebook has created a set of policies, its Community Standards, detailing what type of content people can and cannot post. For instance, Facebook prohibits and removes hate speech, which it defines as ‘content that directly attacks people based on their: race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender, or gender identity, or serious disabilities or diseases’. Although it does not allow hate speech, sometimes people post disagreeable or disturbing content that does not violate Facebook’s policies.

To counter that type of disagreeable or extremist content, Facebook has publicly stated that it considers counter-speech – and the tools that their platform provides to help promote it – to play a critical role. Facebook thinks this is not only a potentially more effective way to tackle this problem, but crucially, is also more likely to succeed in the long run.

Counter-speech is a common, crowd-sourced response to extremism or hateful content. Extreme posts are often met with disagreement, derision, and counter-campaigns. Combating extremism in this way has some advantages: it is faster, more flexible and responsive, capable of dealing with extremism from anywhere and in any language and retains the principle of free and open public spaces for debate. However, the forms counter-speech takes are as varied as the extremism they argue against. It is also likely that it is not always as effective as it could be; and some types of counter-speech could potentially even be counter-productive.

Because of its strong belief in the power of counter-speech and the growing interest in a more rigorous and evidence-led approach to understand it better, Facebook asked Demos to undertake a series of research reports, examining the extent to which different types of counter-speech are produced and shared on Facebook.

This short interim report sets out the summary findings of phase I, which looked at how speech which challenges right-wing populist pages across Europe, is produced and shared. Further reports in this series examine speech and content that challenges extreme Islamist ideology, in the UK and beyond.

It is extremely difficult to create an objective definition of a ‘hate page’ and we do not claim that any of the pages included in this study are hateful, or that the content there is ‘hateful’. Instead, we have focused this study on populist right wing Facebook pages, which are frequently accused of being a place where a high volume of hateful content is posted or shared. We refer to these pages as ‘populist right wing’ pages throughout, and have found that there is a very wide range of content posted and shared there. In determining the counter-speech pages, we tried to

explicitly identify pages for which an important part of their identity was to counteract or respond to what they consider to be hateful groups, pages or content.

We believe it is important that the principle of internet freedom should be maintained; and that it should be a place where people feel they can speak their mind openly and freely. We therefore believe that debate, disagreement, and challenge is nearly always preferable to censorship and removal of content, including when dealing with extreme or radical content, whatever its origin. However, we also believe that this can and should be put on an empirical basis to help us better understand the phenomena and how to respond. This research series is an attempt to do that.

## METHOD

Using Facebook's public 'API' (Application Programming Interface), we collected public posts and interaction data from the public Facebook pages of 150 populist right wing and counter-speech pages from the UK, France, Italy, and Hungary. We used 'R', an open source software that allows researchers to access publicly available data from the public API.

We did not attempt to collect or use any personal information about individuals; nor did we attempt to identify any individuals. Where a user's name or ID was collected inadvertently, it was deleted. We did not collect any data from groups or from individual's pages; and we did not collect any data from closed or secret pages. Throughout, only data from pages that were public and viewable by everyone were used. In order to further protect individual privacy, we have not quoted or republished any specific posts that might identify individuals.

The precise pages were chosen by Demos researchers, and therefore should not be seen as a comprehensive sample of relevant pages. This is not an ideal way of identifying relevant pages, and we stress that this is a pilot project. Further papers in the series develop the methodology further.

Over 2 months (1<sup>st</sup> October 2014 – 1<sup>st</sup> December 2014) we collected 27,886 posts uploaded on these 150 pages from the UK, France, Italy and Hungary. (However, most of the analysis relates to the UK, France and Italy). 'Posts' in this sense refer to updates that were made on the page by the administrator(s) of that page. In addition to posts, we collected all the interactions that were associated with the posts. Interactions refers to 'likes', 'shares' and 'comments' on those posts. Interaction data can be useful in estimating the reach of content, because each time a user interacts with a piece of content, it will appear in their friends' timeline (depending on the privacy settings applied). In total this was 8.4 million interactions.

We subjected this data to a series of analyses. This included: calculating average interactions using automated API results; calculating the format of the most popular types of data using automated API results; calculating the type and style of the most popular types of content through human manual analysis; calculating the types of speech occurring on different pages using human manual analysis ; calculating the type of network within these pages, using automated network analysis ; calculating the way different types of content was shared on pages vis-à-vis users' own newsfeeds using automated analysis.

It is important to stress that these are in many cases quite experimental methodologies. There are no firmly established 'best practice' methods to collect and analyse data of this nature. Further, this is designed as a scoping study. Therefore findings need to be read with caution.

## FINDINGS

### Overall data on the size and scale of hate and counter-speech

In total, we identified 124 populist right wing pages across the four countries, compared to 26 counter-speech pages (for reasons set out above). Similarly, there were many more posts made on populist right wing pages (25,522) than counter-speech pages (2,364). Unsurprisingly, therefore, there were also many more interactions. Populist right wing pages had a total of 7.8 million, compared with 546 thousand for counter-speech pages.

Because of the way the data was collected, our data set includes far more populist right wing pages than counter-speech pages. This does not necessarily mean there are more of them, but more likely that counter-speech pages are more difficult to easily identify. Therefore in order to present a more accurate picture, we calculated the average posts and average interactions per post across each country and each type of page. This found that, overall, in the UK counter-speech pages are smaller in number and more limited in their activity; but achieve a greater amount of sharing and interactions than populist right wing pages.

**Table 1 Average Interactions per post across Countries**

Country	Counter-speech	Populist right wing
France	126	285
Italy	37	235
UK	402	325

### An analysis of network / membership structure and how that effects how content is shared

In order to better understand the way information and ideas flow across these pages, we selected the UK pages and examined the extent to which individuals who commented in one page also commented in another page. This data was taken from Facebook's public API through 'R' and visualized in Gephi, an open source network analysis tool.

With respect to the populist right wing pages, we analysed 92 pages. A total of 54,495 unique users contributed to these pages, who between them made 159,437 comments on those pages over the time period. (This is calculated by collecting the comments and working out the number of unique user IDs that contributed to that data set). It finds that 16.2 per cent of users are active on two or more pages; 1.3 per cent are active in four or more.



We then analysed 21 counter-speech pages. A total of 116,534 unique users contributed to these pages, who between them made 135,842 interactions. Despite having more users these pages have a lower network density: 12.7 per cent are active in two or more counter-speech pages; and 1.15 per cent are active in four or more.

This suggests that the UK based populist right wing pages have a slightly more concentrated network: with fewer, more active users. The counter-speech pages have more active contributors, but do not produce as much content per person. However, this study needs to be replicated at a larger scale to produce significant findings.

### What type of posts are most effective at reaching a wide audience

In order to determine what types of posts were successful in reaching a wide audience, we examined a) the format of posts and b) the content and tone of posts. Facebook API data allows researchers to determine what format posts take, divided by 'link', 'photo', 'status' or 'video'. Table 3 below shows what type of format was most widely used by the pages in question.

**Table 2 Content Types by Country (rounded to nearest per cent)**

	Country	Link	Photo	Status	Video
Counter-speech	France	64	15	14	6
	Hungary	68	5	26	2
	Italy	29	45	17	9
	UK	40	34	18	8
Populist right wing	France	26	43	17	14
	Hungary	10	83	2	5
	Italy	59	27	9	5
	UK	57	23	10	10

This finds a high disparity across countries and pages in terms of the most popular way to share content. For example, in France links were the most popular among counter-speech pages; in Italy it was photos.

The extent to which one type of format is better than another is likely to depend on several factors. However, by calculating the average interactions for each type of format it is possible to draw some conclusions: most principally that photos are the most likely to generate interactions with users (see table 4, below).

**Table 3 Most frequently used Content Type by Country (most total interaction)**

	France	Hungary	Italy	UK
Populist right wing	Photo	Video	Photo	Photo
Counter-speech	Photo	Photo	Photo	Photo

### **Analysis of post content**

In order to have a more nuanced understanding of what type and what tone of post is most popular, we analysed 1000 randomly selected posts from populist right wing pages (split between France, Italy and the UK) and 1000 randomly selected posts from counter-speech pages (again split between France, Italy and the UK). We divided these into categories of ‘content’ of post and ‘tone’ of post. The categories were selected by researchers and are described, with examples, below. We judged post popularity by the interactions they received.

This found that the most popular content of posts on populist right wing pages is what we describe as commentary (2,524 interactions on average).<sup>1</sup> These posts received around double the number of interactions than other content.

The most popular content of posts on counter-speech pages was questions (10,934 interactions on average, although this was skewed by a small number of very popular pieces of content). The second most popular content of posts on counter hate pages was commentary, although in France ‘attacks’ were the most popular category.<sup>2</sup>

The most popular tone of posts on populist right wing pages was ‘celebratory’, such as in posts commemorating war dead or patriotic pride (6,607 interactions on average).<sup>3</sup> ‘Angry’ content was second highest (4,093).<sup>4</sup>

The most popular tone of posts on counter-speech pages across the three countries was funny or satirical tone (2,717).<sup>5</sup>

In order to better analyse the most popular types of posts, we examined 100 posts from counter hate pages that were labelled as satirical, as this was the most successful type of post. This was done in English only.

In terms of interactions, immigration, race and religion all received, on average, over 4,000 interactions and were the most popular types of content. They usually parodied the extremist language used on these issues by hate pages. The most popular piece of content was a request by an army regiment to remove an image celebrating them from ‘Britain First’, the largest populist right wing page in the UK. A screenshot of the request was liked over 9,000 times.

## **An analysis of the type and nature of comments**

Unlike posts, which are published by page admins, anyone can post a comment under a post. We analysed 1000 randomly selected comments from randomly selected posts on populist right wing pages (split between France, Italy and the UK) and 1000 randomly selected comments from randomly selected posts on counter-speech pages (again, split between France, Italy and the UK). The categories were selected by researchers. We judged a comment's popularity by the number of likes they received (this is the only type of interaction available on a comment).

On populist right wing pages, nine per cent of all comments were categorised as counter-speech, meaning comments which disagreed with the post or presented an alternative, more positive message. (This was 17 per cent in France, seven per cent in Italy, and five per cent in the UK). If extrapolated, this would suggest that there are as many as 25 thousand counter-speech comments taking place on populist right wing pages each month in the UK alone.

On counter-speech pages, 'constructive counter-speech' was the most popular successful type of comment (average 5.9 likes per comment).<sup>6</sup> This was followed by 'constructive discussion' (average 5.3 likes per comment).<sup>7</sup> By contrast, 'non-constructive counter-speech' received on average 3.3 likes per comment, and 'fact checks' (3.8 likes per comment).<sup>8</sup> However, despite the popularity of this content, constructive counter-speech accounted for only six per cent of all content, compared to 20 per cent for non-constructive counter-speech.

In order to better understand the specific type of comments that were most popular, we analysed 100 'constructive discussion' comments made in English on counter-speech pages (again, the categories were selected by researchers). This found that 'discussion of broad policy areas' have the most number of average likes (13.2), followed by 'questioning party / movement policy detail' (8.7). This suggests that comments about specific policy areas, policy details, or information is particularly effective at reaching a wide audience. However, these made up a very small proportion of the total number of comments (6 and 3 per cent respectively).

## **Analysis of where and by whom content is being shared and discussed**

We examined the way content spreads from pages to users' own newsfeeds, and whether where content is viewed makes a difference to how that content is interacted with. We analysed the interactions from 4,388 counter-speech posts and 75,132 populist right wing posts, and worked out the proportion that was shared either from the original page or from another user. This is based on a larger data set

than the one used for the research above – although the pages are the same, the data collection period was longer. This data set was only collected with assistance from Facebook’s Data Science team and consisted of public, aggregate, non-experimental, historical data.

**Table 4 Where interactions on posts takes place**

	Populist right wing posts			Counter-speech posts		
	<i>% from original page</i>	<i>% from another user</i>	<i>N=</i>	<i>% from original page</i>	<i>% from another user</i>	<i>N=</i>
<b>Likes</b>	94	6	4,138,425	82	19	56,884
<b>Shares</b>	97	3	474,558	99	1	13,430
<b>Comments</b>	73	27	2,173,678	60	41	16,232

This suggests that as a percentage of total likes, counter-speech pages are better at getting likes and comments on ‘reshares’ – i.e. from people who interacted with content on another user’s newsfeed, rather than on the original page (although they may also like the page where it was originally posted). This means the content can potentially go further. The content could potentially go much further if people would share it more widely with their friends. Further, it is our strong view that there are likely to be a lot of comments / discussion taking place on reshares – i.e. on other people’s newsfeeds, amongst their friends. This, we believe is likely to be where a lot of counter-speech is taking place, although we did not collect this for privacy reasons.

Using the same method, we also examined the way in which different types of content reaches users who did not like the page where it was originally posted (in other words, does content reach beyond those users who have liked a page). We calculated this by working out the percentage of people who liked or commented on posts that had not liked the page where the post was originally posted.

**Table 5 How far content reaches beyond the people who like pages**

	% of likes by people who do not like the page		% of comments by people who do not like the page	
	<i>Populist right wing</i>	<i>Counter-speech</i>	<i>Populist right wing</i>	<i>Counter-speech</i>
<b>Links</b>	50	18	66	32
<b>Photos</b>	52	5	65	25
<b>Videos</b>	68	26	75	48
<b>Status</b>	21	7	41	23

This does show that populist right wing pages are significantly more effective at posting content which goes beyond their network of page fans. For counter-speech pages (and populist right wing pages) videos are the most effective type of content to post to reach a broader audience.

## RECOMMENDATIONS

Based on the above findings, counter-speech pages are not as active as populist right wing pages. In the UK they are doing well in terms of average interactions, but less so in France or Italy. If they wish to reach more people, France and Italian counter-speech pages should produce more content. In the UK, counter-speech pages have more contributors, but they interact less frequently than contributors to populist right wing pages.

For those who wish to encourage the spread of counter-speech, the research also suggests a change in focus would help. Specifically:

- For admins, posting more photos and videos as a proportion of their total output. This should also focus on content that can reach beyond the network of people that like their page.
- For those who comment, more ‘constructive counter-speech’ compared to non-constructive counter-speech; and more comments about specific policy issues.
- Contributors to counter-speech pages should encourage friends to share more content with their friends.
- Overall, this suggests that if counter-speech page administrators and users were more active, and changed their content slightly, it could dramatically increase the reach of their messages.

## WAYS FORWARD

Counter-speech is a more complex phenomenon than it might appear. It is important to see counter-speech as much more than simply disagreeing or confronting a piece of content on a public page. Sometimes it is explicit, such as challenging the views when they appear on a newsfeed, or even actively searching out hateful content and challenging it head on. Other responses are less explicit: some may report the hate speech, block or mute the user or share disagreement in a private message. For others, it is setting up humorous or serious groups to oppose a page or individual. On these pages, we have found some constructive discussion and debate. On other occasions we found less constructive counter-speech, such as simply sending abuse or aggressive threats. Very explicitly, we identified: Constructive counter-speech; non-constructive counter hate; fact checking; constructive discussion, satirical oppositional pages, and serious oppositional pages. Some of this counter-speech might be worth encouraging over other types.

Because there are so many different types of counter-speech, it is difficult to determine what constitutes a successful outcome. There is no current consensus as to what is the key metric, but it can be helpfully broken into three types of metrics (all of which we think can be measured, with varying degrees of difficulty).

### Quantitative metrics

Quantitative analysis of social media content has recently become big business. Hundreds of companies have built thousands of tools attempting to quantify social media influence. Much of this analysis is unsuitable for academic analysis, but many of the central concepts are applicable to this study:

- **Engagement is a measure of user interaction.** For example, it might be the ratio of users who viewed the page and those who signed up, or the number of users who viewed a piece of content and shared or 'liked' it. While this is a blunt instrument, it is a useful proxy of potential reach. We have been able to measure this effectively.
- **Volume and exposure.** How many posts are being produced on the topic, and how many unique users are discussing it? Are there spikes, where the discourse density is higher than usual? This can be done but requires access to more traffic data rather than page specific data.
- **Reach measures the spread of a social media conversation.** How large is the audience? Is hate speech being limited to isolated communities (either by the communities themselves or Facebook's personalisation algorithms)? How often does it spill over into the 'feeds' of users outside of these communities? Reach can be a powerful metric when used in combination with, say, engagement. We found some

useful measures of this in the research.

### **Content metrics**

Content metrics are based on trying to judge the type and nature of content rather than volume. This often requires either manual analysis or sophisticated natural language processing based analysis (both of which are possible – but with varying degrees of accuracy). One of the strengths of these approaches is that they can be scaled up and applied on enormous volumes of content at speed.

- Sentiment analysis is a technique that, given an analytic framework, will allow us to determine if a piece of content can be classed as hate speech, counter-speech or neither. It might also be possible to judge how extreme a piece of content is by processing the language used. We judge this to be difficult, but plausible.
- More specifically, hate content and counter-speech content could be classified using natural language processing into various smaller categories: incitement to offline action; genuine discussions, insults, and so on. We judge this to be difficult but plausible.
- One outstanding question is the extent to which a discussion on Facebook evolves or resolves: does posting counter-speech content have an impact on the remainder of the discussion? We judge this to be very difficult to do using automated techniques, but plausible if done manually with human analysts.

### **Real world metrics**

This refers to whether counter-speech online has an effect on long-term attitudes or behaviour off line. This is extremely difficult to calculate with any degree of precision – just as it is difficult to determine if hateful content online has an effect on attitudes or behaviour. Determining sensible and defensible metrics in this regard would require further research.

This scoping study has also identified some areas where a relatively small amount of further research could improve our understanding of the subject.

First, it is important to create a better way of measuring the total volume of populist right-wing and counter-speech content. It is extremely difficult to produce reliable overall figures of different types of counter-speech without a rigorous and scientific method of objectively determining the pages we found. With an ad hoc method of identification, there is a major risk of misleading data. An automated approach based on the most active and most liked pages across countries and categories would provide a far more robust and reliable measure.



Second, there are new types of content that could be studied in further detail. This would include: the total reach of different types of content (rather than just interactions); examining entire threads of conversations below a post and examining how it changes (are there particular groups or users who are more successful than others in resolving arguments and countering hateful speech?); examining content data for posts shared on individual newsfeed (compared to on the page itself).

Third, it would be useful to understand how Facebook as a network of users, sharing populist right-wing and counter-speech, can and should inform our method. This would involve examining how content spreads. By tracking a single piece of content, for example, we can apply network analysis to see who is sharing it, where they got it from, and where it went next. This could help answer some fundamental questions about how these types of content are spread: what paths do they take? Does different content travel in different ways? How does this change between countries?

Fourth, further research would help to determine what might constitute a sensible and robust measure of offline success. Providing more granular and practical insight might require a series of detailed case studies of online campaigns that have had a major offline impact.

## NOTES

<sup>1</sup> A comment is a 'catch-all' label applied to content that is created or shared by the page that commentates on a situation without necessarily referencing outside sources.

<sup>2</sup> An attack is a particularly aggressive piece of content usually directed at a specific group or individual. An attack can be directed at an ethnic group, a politician or an organization.

<sup>3</sup> Celebratory content is that which celebrates the page or its values.

<sup>4</sup> Angry content can use foul language and call for extreme measures in relation to the content.

<sup>5</sup> This refers to all jokes and satirical content.

<sup>6</sup> Any efforts to have a serious discussion about specific subjects relating to hateful content (xenophobia, immigration)

<sup>7</sup> Refers to any efforts to have a serious discussion about politics / news / general interest issues

<sup>8</sup> Non-constructive counter-speech is anything that challenges hateful content, but in a non-constructive way (ie, attacks an individual, is offensive etc). 'Fact check' is querying or checking a fact or claim made.

## Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

### 1 Definitions

- a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.
- b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.
- c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.
- d 'Original Author' means the individual or entity who created the Work.
- e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.
- f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

### 2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

### 3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

- a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;
- b to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

### 4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

- a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phonorecord of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicense the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.
- b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital filesharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

C If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, You must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

## **5 Representations, Warranties and Disclaimer**

A By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

- i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;
- ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

B except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licensed on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

## **6 Limitation on Liability**

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

## **7 Termination**

A This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

B Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

## **8 Miscellaneous**

A Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

B If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

C No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

D This Licence constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

Counter-speech - which argues, disagrees or presents an opposing view - is a potentially important way to deal with extreme or offensive content online. It is fast, flexible and responsive, capable of dealing with extremism from anywhere, in any language and retains the principle of free and open public spaces for debate. However, it is also likely that it is not always as effective as it could be; and some types of counter-speech could potentially even be counter-productive.

This short interim report sets out the summary findings of new research looking at how speech which challenges right-wing populist pages across Europe, is produced and shared on Facebook. It is based on examining the activity of 150 Facebook pages over a two month period, and how content produced on them is shared and interacted with by users. It also out new methodologies and approaches for measuring how content spreads online, how its impact might be measured, and its effectiveness improved.

This is the first report in a series of research examining counter-speech and content that challenges extreme content online. Future reports examine Islamist ideology in the UK and beyond.

Jamie Bartlett is Director of the Centre for the Analysis of Social Media at Demos. Alex Krasodonski-Jones is a researcher at the Centre for the Analysis of Social Media at Demos.

ISBN 978-1-909037-93-9