

FREE TO BE EXTREME

BY NIKITA MALIK



This video has been removed for violating YouTube's policy on hate speech. Learn more about combating hate speech in your country.

[Learn more](#)



**CENTRE ON
RADICALISATION
& TERRORISM**

Published in 2020 by The Henry Jackson Society

The Henry Jackson Society
Millbank Tower
21-24 Millbank
London SW1P 4QP

Registered charity no. 1140489
Tel: +44 (0)20 7340 4520

www.henryjacksonsociety.org

© The Henry Jackson Society, 2020. All rights reserved.

The views expressed in this publication are those of the author and are not necessarily indicative of those of The Henry Jackson Society or its Trustees.

Title: "FREE TO BE EXTREME"
Author: Nikita Malik

ISBN: 978-1-909035-53-9

£20.00 where sold

Front cover image: Screenshots of YouTube videos filed in an archived collection of online extremist material, Centre on Radicalisation and Terrorism at the Henry Jackson Society, January 2019.

FREE TO BE EXTREME

BY NIKITA MALIK

WITH RESEARCH ASSISTANCE BY CHERMAINE TAY,
DANIEL LEWICKI, ANGELICA GOTTLIEB, AND HANNAH DICKSON



**CENTRE ON
RADICALISATION
& TERRORISM**

About the Author

Nikita Malik is the Director of the Centre on Radicalisation and Terrorism (CRT) at the Henry Jackson Society.

She is an internationally recognised expert on countering violent extremism, terrorism, and hate-based violence, with a focus on youth deradicalisation. In her role, she has worked with key policy makers and government departments in the UK and globally.

A key component of Nikita's work focuses on the propagation of extremist material online, including on social media platforms and the Darknet. Her research has put forward a number of solutions to foster engagement between UK government policymakers and technology companies.

Her past research reports have been regularly featured in the media, and her findings and policy recommendations have been discussed in the Houses of Parliament, European Parliament, the US State Department, and the United Nations.

In addition to working with government policy makers and the media, Nikita has also engaged NGOs to build partnerships and deliver high-impact programmes on youth and radicalisation, including Save the Children, Child Soldiers, Child to Child, UNICEF, and others. Her work on radicalisation has paralleled models on how children are exploited by gangs, cults, and violent groups, and the way technology can help or hinder their exit.

Her work to date led to *Forbes* magazine honouring her in 2018 as a '30 under 30', and a key influencer in law and policy. Nikita was educated at the University of Oxford (where she completed her MA and MSc) and has a second MSc in Middle Eastern Politics and Arabic from SOAS, University of London. She is fluent in four languages.

Acknowledgments

I am extremely grateful for the dedicated research support provided by Henry Jackson Society interns Angelica Gottlieb, Hannah Dickson, Daniel Lewicki, and Chermaine Tay.

I would also like to thank Professor Katharine Gelber (The University of Queensland), Professor Stuart Macdonald (Swansea University), Professor Mark Briers (Alan Turing Institute), Professor Sarah Sorial (Macquarie University), Dr. Erin Saltman, Dr. Andrew Foxall, Coman Kenny, and a number of anonymous reviewers for their helpful comments and feedback.

This research was funded with a research grant gifted by Facebook.

Foreword

In April 2019 I was involved in setting out the Government's plans for a world-leading package of measures to keep UK users safe online. This package consisted of legislative and non-legislative measures to make companies more responsible for their users' safety online, especially children and other vulnerable groups.

One of the most challenging areas technology companies have had to monitor is that of 'extremist' material, often hosted on these platforms for too long, and to the detriment of young and vulnerable people. The 107 cases profiled in this report, comprising of 'extremist-related' offences from 2015-2019, shed light on how individuals are able to manipulate terms and conditions of social media platforms to reach new audiences, gain legitimacy, and raise money.

More than ever, there is need for a regulatory framework benchmarking what is meant by harm, focusing on reducing access to individuals seen as especially harmful. A consistent, cross-platform framework of 'online extremist harm', as put forward in this paper, can be used to assess and flag patterns of behaviour, focusing on both violent and non-violent (harmful) extremism. Technology companies can consider tailored approaches based on similarities of those individuals falling within specific harm categories. Of course, there is space for such a framework to evolve with time and in response to certain offline events, such as terrorist attacks.

As the extremism grading scale evolves, and the behaviour of the individual in question becomes more harmful, it is clear faster action will need to be taken to remove content, users, or organisations from platforms. An independent regulator on online harms, as put forward in the online harms white paper last April, would assist with ensuring that a framework is implemented consistently across platforms, including lesser-known, 'alt-tech' platforms where content and banned individuals and groups may migrate. The indicators in question could be dynamic, evolving as more information becomes available from the offline space.

While some companies have taken steps to improve safety on their platforms, progress has been too slow and inconsistent overall. The report finds that the top platforms used by extremist offenders continue to be big players in the technology industry: Facebook (29.3%), Twitter (14.7%), Whatsapp (14%), Telegram (9%), and YouTube (7.8%), with others making up 25.2% collectively

Our challenge as a society is to help shape an internet that is open and vibrant but also protects its users from harm. Social media platforms play a role in legitimising such speakers (by granting them a blue tick or an account with a large following, for example), providing them with audience amplification techniques (including paid sponsorship and advertisements of posts and events), and allowing their speech to reach new audiences (through algorithms that direct users to similar accounts of interest). This report puts forward an innovative programme of action for multiple actors to assess and respond to such efforts. Technology companies and governments would benefit from taking note of and discussing further its recommendations.

Jeremy Wright QC

*Secretary of State for Digital, Culture, Media and Sport from 2018 to 2019
and co-author of the Online Harms White Paper 2019*

About CRT at The Henry Jackson Society

The Centre on Radicalisation and Terrorism (CRT) is unique in addressing violent and non-violent extremism. By coupling high-quality, in-depth research with targeted and impactful policy recommendations, we aim to combat the threat of radicalisation and terrorism in our society.

The Henry Jackson Society is a think-tank and policy-shaping force that fights for the principles and alliances that keep societies free, working across borders and party lines to combat extremism, advance democracy and human rights, and make a stand in an increasingly uncertain world.

The Henry Jackson Society is a company limited by guarantee registered in England and Wales under company number 0765741 and a charity registered in England and Wales under registered charity number 1140489. HJS is a 501(c)(3) organisation in the United States EIN 46-1190785.

For more information, please see **www.henryjacksonsociety.org**.

Executive Summary

1. Overview

This report explores the balance between freedom of expression and the increasing demands placed on technology companies to monitor harmful extremist content online. In doing so, it sheds light on the evolution of power dynamics between the British government, citizens, and technology companies in defining and enforcing boundaries of acceptable expression and conduct.

There are three overall themes to emerge from the analysis. First, like courts, social media companies face difficulties in applying consistent approaches to banning extremist content, as opposed to monitoring a speaker who is seen as 'extremist'. Second, there continues to be a lack of transparency between companies in communicating information on which organisations and individuals are extremist. Third, courts and social media companies seem to operate in tandem, employing different frameworks to assess harms of extremism. This report advocates for a new, combined, and coordinated framework, where social media companies reference legal cases to assist in understanding extremism, and efforts to monitor content and speakers are overseen by an independent regulator.¹

The framework puts forward a grading scale for online extremist harm. Between 2015 and 2019, 107 cases within the UK where the judge used various offences (terrorism acts, religious and racial hatred, incitement, malicious communication) to police extremism online were examined, using 20 indicators to grade levels of extremism from a 0-15.2 scale. Such a scale allows for different approaches to be employed to react to extremism, as opposed to a binary 'ban or no-ban' option.

1.1. Key Findings and Recommendations

- An online framework can incorporate aspects of case law to set a threshold of indicators of extremism, based on themes such as immediacy of violence, pervasive and potentially offensive or dangerous public pronouncements, and a history of extremist intent.
- 29.3% of the offenders analysed in the report used Facebook as a means to disseminate their extremist views and content (see Figure 1 for more data). However, public forums were often used in conjunction with encrypted applications and 'alt-tech' platforms, including Discord, Surespot, and Telegram. It is therefore important that multiple stakeholders monitor the migration of offenders from one platform to another, where they may take their followers and continue to promote extremist views.
- Despite the use of alternative platforms, individuals disseminating extremist content continued to use mainstream social media platforms due to the legitimacy and influence that such platforms afforded them, as well as opportunities to disseminate views to a wider audience.
- There is a lack of consistency between social media platforms, and discord in approaches to monitoring content within the same company.² Companies should better disseminate information on extremist speakers and content between themselves, perhaps through an independent Global Internet Forum to Counter Terrorism (GIFCT).

¹ An independent regulator was announced in April 2019. 'UK to introduce world first online safety laws', Department for Digital, Culture, Media & Sport, & Home Office, 8 April 2019, available at: <https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws>, last visited 11 December 2019.

² Extremist material could be found in a Google search when not immediately available or blocked on YouTube.

Figure 1 illustrates the data analysis of 107 'extremist-related' cases online (where hate speech, terrorism, or other laws were applied). Statistics to emerge include:

- The majority of hate material shared was Islamist (shared by 71.9% of offenders), followed by material of a far-right nature (shared by 27.1%).
- Terror-related charges were used against 75.5% of cases profiled. This was followed by public order-related offences, used against 22% of cases.
- The largest organisational affiliation of offenders in the dataset was Islamic State (61.2% of cases) followed by no affiliation (anti-Muslim) at 12.1%, and then no affiliation (anti-Semitic or Nazi-related content) at 10.2%. 6 of 107 cases in the dataset (5.6%) were affiliated with Al Qaeda in the Arabian Peninsula.
- Offenders using online platforms for extremist purposes are young, with age at conviction being 15-22 (30.8%) followed by 23-30 (28.9%).
- Of the 107 cases examined, 87% were male and 13% were female.
- 21% of offenders were employed, 7% were unemployed, and data was not known for the majority (65%).
- 56.54% of offenders were active members of a terrorist organisation or extremist group.
- 24.29% of offenders ignored warnings from technology companies, friends, family, or the police.
- 19.15% of offenders had a history of criminal behaviour.

Social media platforms were primarily used for the following reasons:

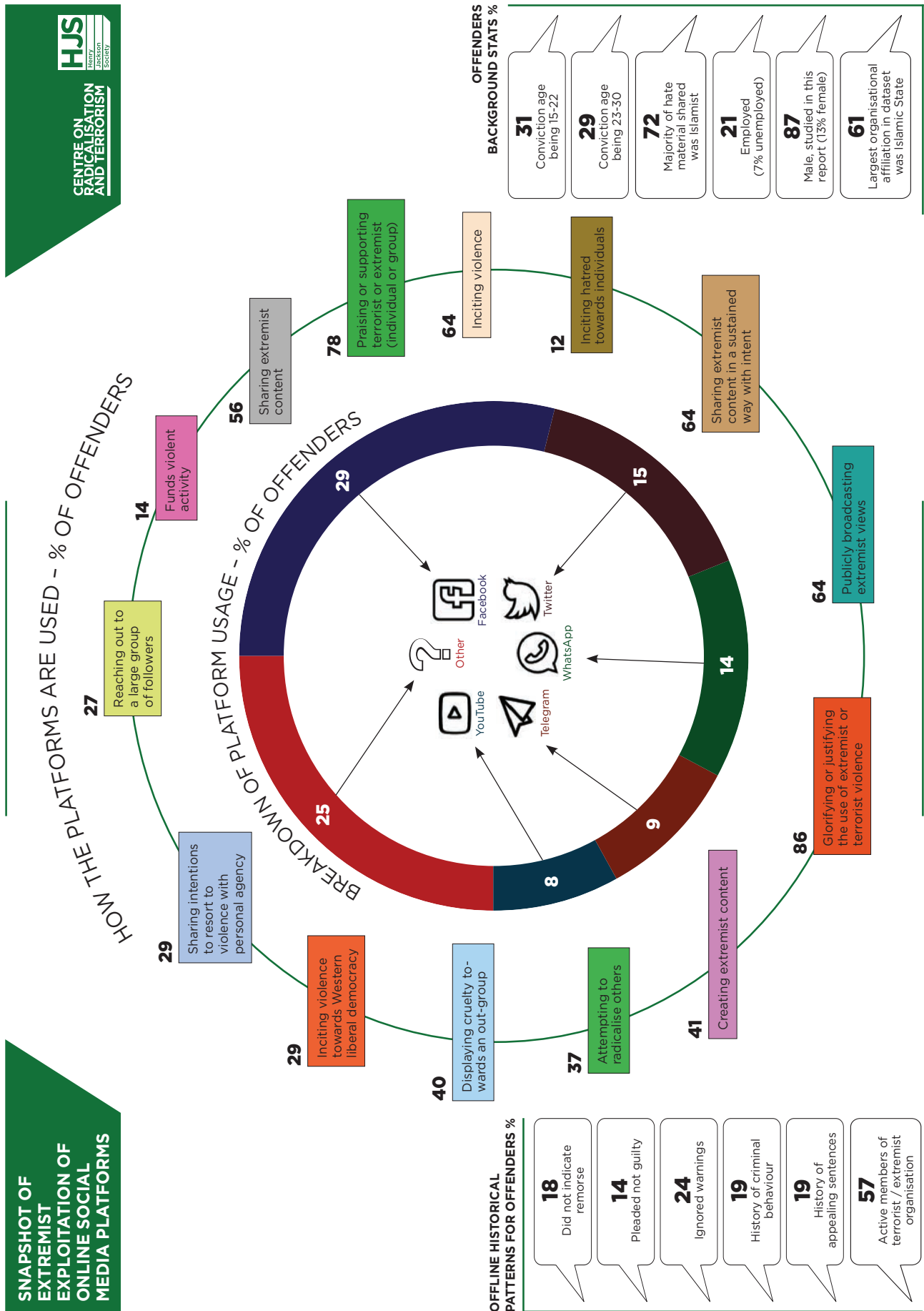
1. Glorifying or justifying the use of extremist or terrorist violence (86.44%)
2. Praising or supporting terrorist or extremist (individual or group) (77.5%)
3. Inciting violence (64.48%)
4. Sharing extremist content in a sustained way with intent (a history of extremist content sharing) (64.01%)
5. Publicly broadcasting extremist views (63.55%)
6. Sharing extremist content (56.07%)
7. Creating extremist content (40.65%)
8. Displaying cruelty towards an out-group (40.18%)
9. Attempting to radicalise others (37.38%)
10. Inciting violence towards Western liberal democracy (29.43%)
11. Sharing intentions to resort to violence with personal agency (28.97%)
12. Reaching out to a large group of followers (27%)
13. Fundraising (13.5%)
14. Inciting hatred towards individuals (11.68%)

The top social media platforms used by offenders were:

1. Facebook (29.3%)
2. Twitter (14.7%)
3. Whatsapp (14%)
4. Telegram (9%)
5. YouTube (7.8%)
6. Other (25.2%)

Offenders tended to use multiple platforms to disseminate and express their views, and mainstream platforms were used in conjunction with 'alt-tech', lesser known platforms (unless the individual was banned from a mainstream platform, which led to a substitution of the mainstream platform by an 'alt-tech' platform).

Figure 1: Graphical Depiction of Findings from Data Analysis



Contents

Foreword	3
Executive Summary	5
1. Overview	5
1.1 Key Findings and Recommendations	5
List of Tables, Figures, and Case Studies	10
Glossary of Abbreviations	11
Glossary of Terms	12
Glossary of Arabic Terms	15
1. Introduction	17
1.1. Background	18
1.2. Defining Concepts	19
1.3. The Spectrum of Extremism.....	21
2. Data Analysis	23
2.1. Extremism Concerns in UK Court Cases	23
2.2. Case Study 1: Zakir Naik	24
2.2.1. Influence.....	24
2.2.2. Extremism Concerns	24
2.2.3. Banned Entry to UK due to Extremism Concerns.....	25
2.2.4. Extremism Definitions and Justifications	26
2.3. Case Study 2: Pamela Geller and Robert Spencer	28
2.3.1. Influence.....	29
2.3.2. Banned Entry to UK due to Extremism Concerns.....	29
2.3.3. Extremism Definition and Justifications	30
2.4. Case Study 3: Raed Salah Mahajna.....	32
2.4.1. Influence.....	32
2.4.2. Extremism Concerns	32
2.4.3. Extremism Definition and Justifications	33
2.5. Case Study 4: AS and FM	34
2.5.1. Influence.....	35
2.5.2. Extremism Concerns	35
2.5.3. Extremism Definition and Justifications	35

2.6. Case Study 5: Alison Chabloz	36
2.6.1. Influence.....	36
2.6.2. Extremism Definition and Justifications.....	37
3. Extremism Indicators: A Framework.....	39
3.1. Categorisation of 20 indicators.....	40
4. The Dataset.....	48
4.1. Limitations	49
5. Extremism Harm Framework: Scaling Harm	50
5.1. Background of Offenders.....	50
5.2. Category Ranking: 0-3.3.....	52
5.2.1. Analysis.....	52
5.2.2. Trends	56
5.2.3. Recommendations.....	56
5.3. Category Ranking: 3.4-5.7.....	57
5.3.1. Analysis.....	57
5.3.2. Trends	59
5.3.3. Recommendations.....	60
5.4. Category Ranking: 5.8-8.1.....	61
5.4.1. Analysis	61
5.4.2. Trends	63
5.4.3. Recommendations.....	64
5.5. Category Ranking: 8.2-10.4.....	65
5.5.1. Analysis.....	65
5.5.2. Trends	67
5.5.3. Recommendations.....	71
5.6. Category Ranking: 10.5-12.8.....	71
5.6.1. Analysis.....	71
5.6.2. Trends	71
5.6.3. Recommendations.....	73
5.7. Category Ranking: 12.9-15.2	73
5.7.1. Analysis.....	73
5.7.2. Trends.....	73
5.7.3. Recommendations	75
6. Broader Recommendations and Conclusions.....	76

List of Tables, Figures, and Case Studies

Figure 1: Graphical Depiction of Findings from Data Analysis	Page 7
Figure 2: The Spectrum of Extremism	Page 22
Case Study 1: Zakir Naik	Page 24
Case Study 2: Pamela Geller and Robert Spencer	Page 28
Case Study 3: Raed Salah Mahajna.....	Page 32
Case Study 4: AS and FM.....	Page 34
Case Study 5: Alison Chabloz.....	Page 36
Figure 3: Indicator Checklist	Page 39
Figure 4: Age at Conviction	Page 50
Figure 5: Geographical Breakdown of Offenders	Page 51
Figure 6: Use of Social Media Platforms by Category Rankings.....	Page 51
Figure 7: Primary Charge against Offenders by Category Rankings	Page 52
Figure 8: Organisational Affiliation by Category Rankings.....	Page 52
Figure 9: Influence Indicators by Category Rankings	Page 53
Figure 10: Average Sentence in Months by Category Rankings	Page 54
Figure 11: History of Criminality and Funding by Category Rankings.....	Page 55
Figure 12: Glorification and Support by Category Rankings.....	Page 58
Figure 13: Interquartile Range Sentence in Months by Category Rankings.....	Page 61
Figure 14: Sharing and Creation by Category Rankings	Page 63
Figure 15: Incitement of Violence by Category Rankings.....	Page 66
Figure 16: Incitement and Cruelty by Category Rankings.....	Page 67
Figure 17: Membership by Category Rankings	Page 68
Figure 18: Appeals Process by Category Rankings.....	Page 70
Figure 19: Use of Social Media Platforms in Category Ranking 10.5-12.8.....	Page 72
Figure 20: Online Extremism Harm Thresholds.....	Page 79

Glossary of Abbreviations

ALM	Al-Muhajiroun
AQAP	Al-Qaeda in the Arabian Peninsula
CAA	Campaign Against Antisemitism
CCE	Commission for Countering Extremism
CPS	Crown Prosecution Service
CST	Community Security Trust
DCMS	Department for Digital, Culture, Media, and Sport
EAU	Extremism Analysis Unit
ECHR or the Convention	European Convention on Human Rights
EDL	English Defence League
FGM	Female Genital Mutilation
FOI	Freedom of Information
FtT	First-tier Tribunal
GIFCT	Global Internet Forum to Counter Terrorism
HBV	Honour-based violence
IRF	Islamic Research Foundation
IS	Islamic State
LeT	Lashkar-e-Taiba
LGBTQI+	Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, Intersex+
MEMO	Middle East Monitor
NA	National Action
OSCT	Office for Security and Counterterrorism
PGP	Pretty Good Privacy
PKK	Partiya Karkeren Kurdistanî
TA	Terrorism Act
Tor	The Onion Router
UAF	Unite Against Fascism
UK	United Kingdom
US	United States
UtT	Upper-tier Tribunal
VK	VKontakte

Glossary of Terms³

Ahrar al-Sham, includes Harakat Ahrar al-Sham al-Islamiyya

A Sunni Salafist militant group operating in Syria that aims to replace the Assad Regime with an Islamic government.

Al Ghurabaa, includes The Saved Sect, al-Muhajiroun, Islam4UK, Call to Submission, Islamic Path, London School of Sharia, and Muslims Against Crusades

(Proscribed July 2006) An Islamist group which seeks to establish an Islamic *caliphate* ruled by *sharia* law. The group first emerged as al-Muhajiroun in the UK, in 1996, led by Omar Bakri Muhammed. While the Group has some links to groups overseas, it is based and operates within the UK.

Al-Qaeda (AQ), includes Al-Qaeda in Iraq (AQI), Al-Qaeda in the Arabian Peninsula (AQAP), and Al-Qaeda in the Islamic Maghreb

(Proscribed March 2001) Inspired by Osama bin Laden, the group's aims include the expulsion of Western forces from Saudi Arabia, the destruction of Israel, and the end of Western influence in the Muslim world.

Ansar Al-Sharia, includes Ansar al-Sharia-Benghazi (AAS-B)

(Proscribed November 2014) A Sunni Islamist militia group involved in terrorist attacks against civilian targets and frequent assassinations and attempted assassinations of security officials and political actors in eastern Libya. AAS-B continues to pose a threat to Libya and Western interests and is alleged to have links to proscribed organisation Ansar al-Sharia-Tunisia and AQ.

Appeal

The process by which parties to cases seek a re-examination of a case.

Discord

A platform designed for video gaming communities, specialising in text, image, video, and audio communication between users in a chat channel.

Dog whistle

Political messaging employing coded language that appears to mean one thing to the general population but has an additional, different, or more specific resonance for a targeted subgroup.

English Defence League (EDL)

A far-right pressure group and social movement that claims to be opposing Islamic extremism. Its rhetoric frequently targets Islam and Muslims in general.

Extremism

An ideology which, when implemented, would significantly and negatively impact the human rights of certain sectors of society, such as women, religious or ethnic groups, or persons with disabilities. By extension, violent extremism is an ideology that would justify the use of violence against these sectors of society.

Freedom Defense Initiative, includes Stop Islamization of America (SIOA)

An American organisation known primarily for its controversial anti-Muslim advertising campaigns.

Gab

An English-language social media network which aims to promote free speech.

³ The proscriptions used herein have been issued by the Home Office, published in 'Proscribed terrorist groups of organisations' (2013), available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/795457/Proscription_website.pdf, last visited: 16 October 2019.

Hamas Izz al-Din al-Qassem Brigades, includes Hamas

(Proscribed March 2001) Hamas aims to end Israeli occupation in Palestine and establish an Islamic state.

Hate Crime

A term used to describe a range of criminal behaviour where the perpetrator is motivated by hostility or demonstrates hostility towards the victim's disability, race, religion, sexual orientation or transgender identity. A hate crime can include verbal abuse, intimidation, threats, harassment, assault, and bullying, as well as damage to property.

Indoctrination

To teach a specific viewpoint or ideology without allowing anyone to criticise or question it, often in reference to religious ideas.

Islamic Movement in Israel, includes Islamic Movement in 48 Palestine

An Islamist movement that aims to advocate Islam in Israel, particularly among Arabs and Circassians.

Islamic State (IS), includes Islamic State of Iraq and the Levant (ISIL), Dawlat al-Iraq al-Islamiyya, Islamic State of Iraq (ISI), Islamic State of Iraq and Syria (ISIS) and Dawlat al-Islamiya fi Iraq wa al Sham (Daesh)

(Proscribed June 2014) A brutal Sunni Islamist terrorist group active in Iraq and Syria. The group adheres to a global jihadist ideology, following an extreme interpretation of Islam, which is anti-Western and promotes sectarian violence. IS aims to establish an Islamic state governed by *sharia* law in the region and impose its rule on people using violence and extortion. IS was previously proscribed as part of AQ.

Islamism

The belief that Islam is a totalitarian political ideology. It claims that political sovereignty belongs to God rather than the people. Islamists believe that their reading of *sharia* should be state law, and that it is the religious duty of all Muslims to work towards and pledge allegiance to an Islamic state that reflects these principles.

Jabhat al-Nusrah (al-Nusrah Front)

Following an order by the UK in July 2013, the al-Nusrah Front and Jabhat al-Nusrah li-ahl Sham should be treated as alternative names for the organisation that is already proscribed as Al-Qaeda.

Jihadism

A militant strand of Sunni Islamism which advocates the use of violence against non-Muslims (or other Muslim groups such as Shia or Sufi Islam) as part of a broader struggle for the establishment of an Islamic state.

Kateeba al-Kawthar (KaK), includes Kataib al-Muhajireen, 'Ajnad al-sham' and 'Junud ar-Rahman al Muhajireen'

(Proscribed June 2014) KaK is an armed terrorist group fighting to establish an Islamic state in Syria. The group is aligned to the most extreme groups operating in Syria and has been linked to Al-Qaeda. KaK is believed to attract a number of Western foreign fighters and has released YouTube footage encouraging travel to Syria and asking Muslims to support the fighters.

Keyboard warrior

A person who makes abusive or aggressive posts online, typically concealing their true identity.

Lashkar-e-Taiba (LeT), includes Lashkar e Tayyaba (LT)

(Proscribed March 2001) The group seeks independence for Kashmir and the creation of an Islamic state using violent means.

London Forum

Emerging in 2011 following a split from the New Right, the group is described as one of the organising hubs for the far-right in the UK.

National Action, includes Scottish Dawn, NS131 (National Socialist Anti-Capitalist Action)

(Proscribed December 2016) A racist neo-Nazi group established in 2013. Its activities and propaganda materials are particularly aimed at recruiting young people. The group is virulently racist, anti-Semitic, and homophobic. National Action's online propaganda material, disseminated via social media, frequently features extremely violent imagery and language.

Paltalk

A video group chat service that enables users to communicate via video, internet chat, and voice. An alternative to the more commonly used platform, Skype.

Partiya Karkeren Kurdistan (PKK), includes Kurdistan Worker's Party, KADEK, and Kongra Gele Kurdistan

(Proscribed March 2001) Primarily a separatist movement that seeks an independent Kurdish state in southeast Turkey.

Pretty Good Privacy (PGP)

An encryption program that provides cryptographic privacy and authentication for data communication. PGP is used for security purposes, such as signing, encrypting, and decrypting texts, emails, files, and other communications.

Proceedings

The series of hearings and other sessions in court that make up the composite parts of a legal case.

Radicalisation

The process by which individuals and/or groups come to adopt extremist ideologies.

Surespot

An encrypted chat messenger with a focus on privacy and security.

System Resistance Network (SRN)

A far-right organisation explicit about its neo-Nazi beliefs. Claims that SRN is run by National Action have not yet been reflected in the UK's proscribed terrorist organisations.

Telegram

A cloud-based instant messaging and voice over service. An alternative to the more commonly used platform, WhatsApp.

Terrorism

The use of violence or illegal force targeted at civilians by non-state actors that seeks to bring about political or societal change.

The Onion Router (Tor)

A free and open-source software for enabling anonymous communication.

Unite Against Fascism (UAF)

An anti-fascist pressure group in the UK.

Vkontakte (VK)

A Russian online social media and social networking service based in Saint Petersburg.

Weebly

A web hosting service specifically oriented for online shopping.

Zintan Brigades, Government of National Accord affiliation

Government of National Accord-funded armed units linked to the town of Zintan in Libya and its surrounding area. The Zintan Brigades are allied to, but separate from, the Libyan National Army.

Glossary of Arabic Terms⁴

Dabiq

Online magazine used by Islamic State for the purposes of radicalisation and recruitment.

Dar al-Harb

Literally translates to 'Lands of War'; a reference to territories outside of the *Dar al-Islam* (Land of Islam), or an Islamic state. An ancient Islamic concept dating back to long before the emergence of contemporary *jihadism*, it is frequently used by Islamist extremists in reference to lands outside of the Islamic *caliphate* such as Europe or North America. Sometimes referred to by Islamists as *Dar al-kufr* (see below).

Dar al-Islam

'Land of Islam'; Islamists commonly define *Dar al-Islam* as any land under Muslim control which implements the religious principles of *sharia* as divine law.

Dar al-kufr

'Land of disbelief'

Dawah / da'wa

Literally translates as 'invitation', the proselytising or preaching of Islam.

Istishhad

The act of deliberately killing oneself with the intent of seeking martyrdom.

Jihad

Literally translates as 'struggle'; interpretations range from a personal effort to live according to Islam to defending Islam by means of an armed struggle, and physically fighting in the way of Allah in order to establish Islam. In the context of this paper (unless stated otherwise), *jihad* should be taken to mean 'armed struggle'.

Kafir (pl. **kaffir** or **kuffar**)

'Non-believer' (referring to non-Muslims); the term could also be used derogatorily to suggest a (Muslim or non-Muslim) person's disbelief in God and/or denial of truth.

Khalifa / Caliph

The ruler of a caliphate.

Khilafa / Caliphate

Islamic state; an expansionist state governed by a *khalifa* and implementing *sharia* as state law.

Kufr

Disbelief

Mujahid (pl. **mujahedeen** / **mujahidin**)

A person who takes part in *jihad* as armed struggle.

Nasheed (pl. **anashid**)

Literally translates as 'chants'; this is an Islamic hymn that is performed acapella or accompanied by percussion instruments. The content usually comprises references to Islamic beliefs and history.

Salafi

Salafists are ultra-conservative followers of Islam. The word *Salafi* was originally used to describe the earliest generations of Muslims, and was later adopted by a modern group of

⁴ Arabic terms have been adapted from Bewley, A., *Glossary of Islamic Terms* (London: TaHa Publishers, 1998).

Muslims led by al-Afghani and Muhammad Abduh at the turn of the century as a response to European imperialism.

Sharia / Shariah

Literally translates as 'road'; the Muslim religious code of conduct; a range of diverse traditions and interpretations of Islamic jurisprudence, from strict rules to broad principles and objectives.

Ummah

The fraternity of believers, the transnational Muslim community.

1. Introduction

No doubt lessons can and have been learnt by many people from the unique circumstances of [RXG's] case but there is no material before me from which blame should be attributed to anyone, except those extremists who were prepared to use the internet to encourage extreme views in a boy of 14 and then use him to carry out terrorist acts.

Justice Nicklin on RXG (aka Boy X), the youngest person to be convicted of a terrorist offence in Britain ⁵

Social media platforms continue to grapple with extremist speakers using their tools to reach new and existing audiences. Policies on the part of these companies have gradually shifted, from an approach of purposefully doing very little in the interests of protecting free speech, to monitoring and removing content that risks exhibiting harms described in their terms and services. New approaches have been varied and piecemeal, resulting in different decisions being made for individuals and organisations who exhibit similar extremist behaviours.

Vague international points of reference on the definitions of terrorism and extremism have contributed to the struggle that technology companies face in moderating content seen as extremist – that is, content assumed to facilitate radicalisation. It is therefore crucial to understand how multiple stakeholders must balance the policing of content online in the interests of protecting the public, while still upholding the tolerance inherent within liberal societies that allows individuals to exercise their rights of freedom of speech and expression. ⁶

This report explores this balance in various ways. Broadly, it aims to develop an account of 'extremism' by creating 20 indicators to assist social media companies, given they have to operate in a definitional vacuum created by lack of legal regulation. The resulting grading system can also be used by government agencies and courts in assessing the permissibility of extremist speech and the level of access extremist speakers can have to audiences.

As such, the report begins with a systematic review of court cases in the United Kingdom (UK) where extremists have appealed decisions made by the Home Secretary on the grounds of violation of the right to freedom of expression pursuant to Article 10 of the European Convention on Human Rights (ECHR or "the Convention").⁷ The five resulting cases are examined in detail, to reveal that courts rely on a number of factors to determine whether a speaker is, in fact, an extremist, including, but not limited to: the gravity and context of the words expressed; the longevity, applicability, and territoriality of these statements; and the level of harm present in the words, either by facilitating, inciting, or causing violence, or by targeting a protected group. Factors weighed when determining whether to ban extremist speakers include the individual's previous history, their level of influence, their perceived target, their intent, the space and context in which the harmful speech occurs, and any expressed remorse. In four of these cases, limitations placed on extremists occur offline; it is the final case examined, that of Alison Chabloz, where measures are taken to reduce an individual's ability to broadcast their views online.

These case studies illustrate that factors used by the courts can – and should – form the basis of a framework to moderate extremism in the online space. After all, social media platforms

⁵ RXG v Ministry of Justice and Ors [2019] EWHC 2026 (QB), available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2019/2026.html&query=\(anzac\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/QB/2019/2026.html&query=(anzac)), last visited: 3 October 2019, p. 1.

⁶ Francois-Revel, J., *How Democracies Perish* (New York: Doubleday, 1983): p. 4.

⁷ It is important to note that the UK does not have a law against 'extremism', and the cases examined in this report comprise of a number of 'extremism-related' offences including hate crime, terrorism, and others. Similarly, technology companies do not regulate against extremism per se, but use multiple offences that may violate their terms and conditions. For more, see Section 5.

play a role in legitimising such speakers (by granting them a blue tick or an account with a large following, for example), providing them with audience amplification techniques (including paid sponsorship and advertisements of posts and events), and allowing their speech to reach new audiences (through algorithms that direct users to similar accounts of interest). Facebook, for example, was used most frequently by 39 out of the 107 extremist offenders in the database compiled in the analysis section of the report (see Section 5, The Dataset). Despite this use of the platform, a great deal of extremist material is available on Telegram, Discord, Weebly, Surespot, and others. This paper finds that lesser-known platforms are often used in conjunction with public and easily available platforms such as Facebook. As a result, the recommendations section of the report advocates for a collaborative approach led by an independent GIFCT to ensure extremist material, individuals, and organisations are not able to disseminate information to larger audiences.⁸

The paper puts forward a system of measurement of extremist or radical material so that Article 10 issues can be dealt with more reliably. Using a grading system of extremist harm online, this report proposes a framework for technology companies and others to react to content based on the historical and contextual background of the user in question, and advocates for the creation of multiple options to respond to extremist individuals and organisations (as opposed to a binary ‘ban or no-ban’ option). The framework seeks to ensure both that platforms are not continuously exploited by extremists and that free speech is protected, particularly for those who criticise extremist content and religion, or use satire, irony, or art to do so.

1.1. Background

The theory of extremism is premised on the fact that those who subscribe to and advocate for such views operate outside the boundaries of toleration. That they are able to do so in a liberal society sheds light on the modern-day paradox of toleration, a paradox often amplified by social media platforms where such views are given greater exposure to international audiences.

Certain academics, such as the noted legal scholar Lee C. Bollinger, have argued that the concept of free speech has expanded to protect speech devoid of any positive value.⁹ For Bollinger, there is no better example of the foundation of tolerance than protecting speech that is ‘extremist’, and doing so is seen as exercising constructive toleration and self-restraint on the part of the listener.¹⁰ On the other side of the debate are those such as Sarah Sorial, professor in law researching the limits of free speech and deliberative democracy, who makes the case that speech which does not add social value does not deserve protection, as the content of one’s speech cannot be separated from one’s goals.¹¹ Any framework attempting to regulate speech that is either extremist in content, spoken by an extremist, or both, will need to balance toleration with the propensity to harm with one’s words.

According to John Stuart Mill’s Harm Principle, “the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others”.¹² Therefore, if speech is sufficiently harmful to people’s overall liberty, the argument can be made that it should be regulated. Mill’s Principle proves useful in differentiating harm from mere adversity. Defining ‘harms’ as negative outcomes which are sufficiently detrimental

⁸ The GIFCT is the Global Internet Forum to Counter Terrorism, a group of companies dedicated to disrupting terrorist abuse of members’ digital platforms. For more, see: ‘Global Internet Forum to Counter Terrorism: Evolving an Institution’, Global Internet Forum to Counter Terrorism, available at: <https://www.gifct.org/about/>, last visited: 11 December 2019.

⁹ Rosenfeld, M., ‘Review: Extremist Speech and the Paradox of Tolerance’, *Harvard Law Review* 100.6 (1987): pp.1457-1481.

¹⁰ Bollinger, L., *The Tolerant Society: Freedom of Speech and Extremist Speech in America* (New York: Oxford University Press, 1986): p. 182.

¹¹ Sorial, S., ‘Can Saying Something Make It So? The Nature of Seditious Harm’, *Law and Philosophy* 29.3 (2010): pp.273-305; p. 287.

¹² Mill, J.S., *On Liberty* (New York: Bartleby, 1999): p. 17.

to people's liberty, it provides a justification for state intervention based on the purpose of the prevention of liberty-diminishing outcomes; yet still acts in the interests of people's liberty overall.¹³ Accordingly, speech that is merely bothersome is not sufficiently detrimental to the victim's liberty, and should not be liable to legal restriction. As Carl Cohen, professor on morality and ethics, puts it, "some public offensiveness, and some private distress, will be an inevitable cost of freedom. That is what a free, democratic society requires."¹⁴

As such, in European societies – including the UK – free speech laws have tended to incorporate both the protection of minorities and security considerations as boundaries to what can be said. The Public Order Act 1986 prohibits the use of "threatening, abusive, or insulting words or behaviour" with the intent to cause another to believe that immediate violence would be used or to provoke immediate violence,¹⁵ allowing officials to put restrictions on public processions that aim to intimidate others,¹⁶ and proscribes the use of language and distribution of material that intentionally promotes racial hatred,¹⁷ religious hatred, and hatred on the grounds of sexual orientation.¹⁸ The Malicious Communications Act 1988 further prohibits indecent, grossly offensive, threatening, or false communications.¹⁹ Moreover, regulation is enforced by Article 10 of the ECHR, which argues that free speech can be limited by law to protect the conditions necessary for a democratic society.²⁰ Restricting expression may be required to fulfil the aim of governments and institutions to balance toleration, protection, and freedom in liberal democratic societies. As evidenced, the jurisdiction under which freedom of expression is exercised determines when a speaker may be at fault in abusing this liberty – the law thereby defines the degree of toleration.

1.2. Defining Concepts

The concept of extremism has slowly evolved in public space within the UK. Following the 7/7 bombings in 2005, then Home Secretary Charles Clarke outlined a list of "unacceptable behaviours" that could be used to exclude extremists from the country, however, the term 'extremism' was not defined on this list, and was used only once.²¹ Other plans by the Home Office, including the creation of a clause to explain which views were considered extremist and at odds with a culture of tolerance, were abandoned due to controversy,²² as was the idea of producing a database listing those who violated the unacceptable behaviours policy, and circulating this list amongst British immigration officials.²³

¹³ Simpson, R., 'Harm and Responsibility in Hate Speech', University of Oxford (2013), p. 76.

¹⁴ Cohen, C., 'Free Speech and Political Extremism: How Nasty are We Free to Be?', *Law and Philosophy* 7.3 (1988-1989), p. 277.

¹⁵ Public Order Act 1986, Chapter 64, Part I, 4, available at: www.legislation.gov.uk/ukpga/1986/64/contents, last visited: 12 July 2019.

¹⁶ Public Order Act 1986, Chapter 64, Part II, 14.

¹⁷ Public Order Act 1986, Chapter 64, Part III, 18. Freedom of Expression was not legally codified in the UK until the Human Rights Act of 1998, which incorporates the ECHR into British law. Article 10 of the Convention guarantees everyone's right "to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers." See Human Rights Act 1998, Schedule 1, Article 10, available at: <http://www.legislation.gov.uk/ukpga/1998/42/contents>, last visited: 20 August 2019.

¹⁸ Public Order Act 1986, Chapter 64, Part 3A, 29C.

¹⁹ Malicious Communications Acts 1988, Chapter 27, Article 1, available at <https://www.legislation.gov.uk/ukpga/1988/27/contents>, last visited: 12 July 2019.

²⁰ The right to freedom of expression can be restricted "in the interests of national security, territorial integrity or public safety, for the prevention of disorder and crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary". See European Convention on Human Rights, Article 10.

²¹ Home Office, Departmental Report 2004-2005 (2005), available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/272108/6528.pdf, last visited: 21 October 2019.

²² Tempest, M., 'Clarke reveals terror deportation rules', *The Guardian*, 24 August 2004, available at: <https://www.theguardian.com/world/2005/aug/24/terrorism.uk>, last visited: 19 September 2019.

²³ *Ibid.*

The UK's Channel Duty Guidance defines extremism as the "vocal or active opposition to fundamental British values, including democracy, the rule of law, individual liberty and mutual respect and tolerance of different faiths and beliefs. We also include in our definition of extremism calls for deaths of members of our armed forces, whether in this country or overseas".²⁴ This definition adapted the earlier unacceptable behaviours policy, but placed emphasis on vocal opposition and incitement to violence (as opposed to an actual act of violence) and defined extremism in terms of British values.²⁵ While the existence of non-violent extremism was acknowledged, then Home Secretary Theresa May stated that all extremism eventually leads to violence, including terrorism.²⁶ Examples of those who may hold extremist views included those who reject Western and British values as fundamentally incompatible with Islam, those who promote the establishment of a *caliphate* or state governed according to Islamic *sharia* law, and those who view all those who disagree with them as *kuffar* (disbelievers).²⁷ Individuals involved in the Trojan Horse plot,²⁸ hate preachers, those involved in hate crimes or honour-based violence (HBV), and British citizens travelling to fight in Syria and Iraq were all further identified as examples of extremists.²⁹ However, despite these listed examples of extremism and extremists, the definition was still met with a fair share of criticism for being unclear and inapplicable, leading to the creation of a Counter Extremism Commission in early 2018 to redefine 'extremism'.

A key turning point in naming particular individuals as extremist occurred in 2015 with the publication of the press release 'PM's Extremism Task force: tackling extremism in universities and colleges top of the agenda'. The press release defined extremists as speakers known to have used rhetoric that sought to undermine, or who were at odds with, core British values, much in the same manner as in the Home Secretary's speech on extremism the same year.³⁰ The press release stated that the Home Office's newly established Extremism Analysis Unit (EAU) had identified extremists involved in 70 events held on university campuses in 2014, and, for the first time, explicitly named six speakers, including Salman Butt, as extremists who needed to be subjected to risk assessments by universities at which they wished to speak, and who would not be allowed to speak unchallenged on campuses.³¹ This established the unique responsibility of certain spaces – such as universities – in protecting audiences from extremist speakers.

In response to publication of the press release and being named as an extremist, in October 2016 Salman Butt issued proceedings for judicial review, now concluded, and for libel, which were continuing at the date of publication.³² The judicial review claim was dismissed by the

²⁴ 'Channel Duty Guidance: Protecting vulnerable people from being drawn into terrorism', HM Government (2015), available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425189/Channel_Duty_Guidance_April_2015.pdf, last visited: 5 September 2019, p. 3.

²⁵ 'British values' are defined, for example, by the Department of Education in the UK as the values of democracy, the rule of law, individual liberty, and mutual respect and tolerance of those with different faiths and beliefs. For more, see: 'Guidance on promoting British values in schools published', Department for Education, 27 November 2014, available at: <https://www.gov.uk/government/news/guidance-on-promoting-british-values-in-schools-published>, last visited: 11 December 2019.

²⁶ 'Channel Duty Guidance: Protecting vulnerable people from being drawn into terrorism', HM Government (2015).

²⁷ Ibid.

²⁸ An attempt to promote Islamist values in schools in Birmingham in late 2013. For more, see: 'Trojan Horse 'plot' schools timeline', *BBC News*, 16 July 2015, available at: <https://www.bbc.co.uk/news/uk-england-birmingham-28370552>, last visited: 11 December 2019.

²⁹ 'Channel Duty Guidance: Protecting vulnerable people from being drawn into terrorism', HM Government (2015).

³⁰ Lord Justice Underhill, Lady Justice Sharp and Sir Rupert Jackson, 'Between: Dr Salman Butt and the Secretary of State for the Home Department', 6 June 2019, available at: <https://www.5rb.com/wp-content/uploads/2019/06/Butt-v-SSHD.pdf>, last visited: 19 September 2019; 'PM's Extremism Taskforce: tackling extremism in universities and colleges top of the agenda', *GOV.UK*, 17 September 2015, available at: <https://www.gov.uk/government/news/pms-extremism-taskforce-tackling-extremism-in-universities-and-colleges-top-of-the-agenda>, last visited: 19 September 2019.

³¹ 'PM's Extremism Taskforce: tackling extremism in universities and colleges top of the agenda', *GOV.UK*, 17 September 2015.

³² Lord Justice Underhill, Lady Justice Sharp and Sir Rupert Jackson, 'Between: Dr Salman Butt and the Secretary of State for the Home Department', 6 June 2019.

High Court in July 2017. However, in March 2019, the Court of Appeal overturned that decision and found that the Secretary of State had breached his duty to promulgate guidance that was sufficiently balanced and accurate to inform the decision maker in a university or educational setting of their competing obligations to ensure free speech and prevent people from being drawn into terrorism, and to assist them to a proper conclusion.³³ This decision would mean that the guidance would need to be redrafted by government and approved by Parliament.³⁴ This example illustrates the difficulties in naming an individual or an organisation as ‘extremist’, while still protecting audiences from extremism (for more, see the case of Zakir Naik to follow). It further reflects the need for a framework for extremism that is consistent across cases.

1.3. The Spectrum of Extremism

The lack of a comprehensive and granular definition of extremism partially explains why technology companies have had to rely on their own policies for justifying the removal of extremist content. While extremism can fit into the hate speech category,³⁵ it usually requires a unique approach given that such content dehumanises those groups seen as ‘others’, makes statements on their inferiority, or calls for their exclusion and/or segregation, all of which are often linked to violence. Given the possibility of security threats stemming from extremist activity, faster decisions have to be made than those on hate speech alone.

In the offline space, the Government’s counter-extremism strategy, published in October 2015, identified terrorism and hate crime as the results of extremism.³⁶ It differentiated between violent and non-violent forms of extremism, and listed justifying (even without explicitly promoting) violence, inciting hatred, promoting the operation of alternative systems of law in Britain, the discouragement of participation in the democratic system, and the carrying out of illegal cultural practices such as female genital mutilation (FGM) as forms of extremist activity.³⁷ Given that ‘extremism’ covers both terrorism and hate crime, one can argue that a spectrum exists, where more serious extremist offences advocate for violence (see Figure 2 below).³⁸

When it comes to extremism, the obligation to police has meant that social media companies have gone beyond what is legally required and “forged their own space of responsibility”.³⁹ This is certainly the case with Facebook, which has created its own definition of terrorism,⁴⁰ and

³³ Brick Court Chambers, ‘Court of Appeal declares Prevent Duty Guidance to be unlawful’, 8 March 2019, available at: <https://www.brickcourt.co.uk/news/detail/court-of-appeal-declares-prevent-duty-guidance-to-be-unlawful>, last visited: 16 October 2019.

³⁴ It is important to note that Butt challenged two decisions, the first regarding the lawfulness of the decision to name him as an extremist because this decision interfered with the rights of the speaker and students to impart and receive information, in breach of their rights of free speech; and the collection, storage, use, and dissemination of his personal data by the Extremism Analysis Unit (EAU) which breached his rights of private life and data protection. The latter part of the claim was rejected by both the High Court and the Court of Appeal; Butt is seeking permission to appeal this aspect of his case to the Supreme Court. He also issued proceedings for defamation; One Brick Court report mistakenly stated that Butt’s libel claim had been dismissed.

³⁵ ‘Hate Speech’, Facebook, undated, available at: www.facebook.com/communitystandards/hate_speech, last visited: 5 September 2018.

³⁶ ‘Counter-Extremism Strategy’, HM Government (2015), available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470088/51859_Cm9148_Accessible.pdf, last visited: 19 September 2019, pp.5-6.

³⁷ *Ibid*, pp.10-13.

³⁸ David Parnham, named by the Crown Prosecution Service (CPS) as an “anti-Muslim extremist”, was prosecuted under the Anti-Terrorism Crime and Security Act, amongst others.

³⁹ Grimmelmann, J., ‘Some scepticism about search neutrality’, TechFreedom (2011), available at: <https://nissenbaum.tech.cornell.edu/papers/The-Next-Digital-Decade-Essays-on-the-Future-of-the-Internet.pdf>, last visited: 20 August 2019, pp.435-460. Tushnet, R., ‘User-generated discontent: Transformation in practice’, *Columbia Journal of Law & the Arts* 31.101 (2008): p.497.

⁴⁰ Hollister, S., ‘Here’s How Facebook defines terrorism – and how it’s responding’, *CNET*, 23 April 2018, available at: <https://www.cnet.com/news/facebook-shares-terrorism-definition-al-qaeda-isis/>, last visited: 19 January 2019.

Figure 2: The Spectrum of Extremism

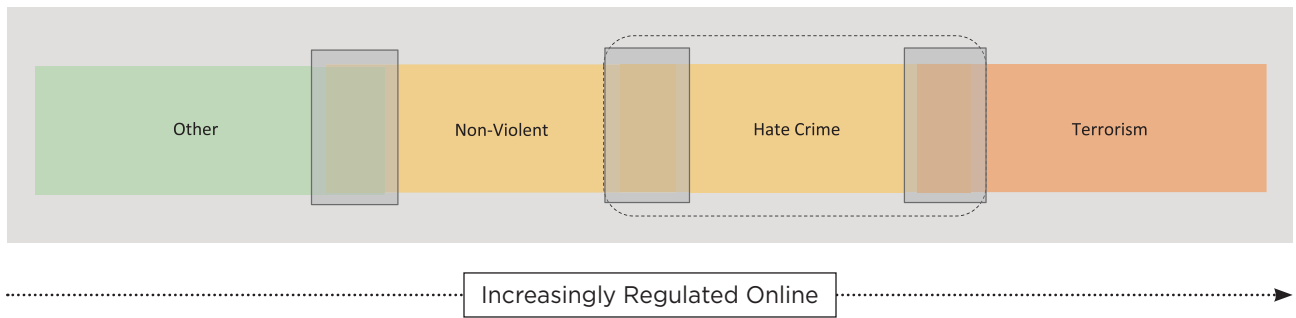


Figure is author's own work

Google, which has attempted to define extremism.⁴¹ Moreover, as these technology companies operate global platforms, national laws on extremism and terrorism make the issue more contentious, as definitions vary from state to state.⁴²

One of the clear issues to emerge from efforts to regulate extremism is the distinction between content that is extreme, and speakers who are extremist. This distinction has yet to be made concretely by the British government, due to various risks of libel and controversy, as per the case of the list of speakers to be banned for unacceptable or extremist behaviour, and the case of Salman Butt above. Based on the five qualitative profiles of those defined as 'extremist' to follow, several patterns of behaviour emerge that can assist in understanding which actions can be taken to monitor and protect against extremism online. These include the history of the speaker, the content of their words, their access to an audience, and the space in which such words are said. These trends form the basis of 20 indicators to follow, that can be used to map an individual's level of extremist harm online, leading to a grading scale with alternative options to restrict a user's speech and the ability for an audience to access this speech. This framework can be employed by multiple agencies to better monitor and understand extremism online (see Recommendations).

⁴¹ 'Featured Policies: Violent Extremism', *Google Transparency Report*, undated, available at: <https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism>, last visited 10 January 2019.

⁴² Harris-Hogan, S., Barrelle, K. and Zammit, A., 'What is countering violent extremism? Exploring CVE policy and practice in Australia', *Behavioural Sciences of Terrorism and Political Aggression* 8.1 (2016): pp.6-24.

2. Data Analysis

2.1. Extremism Concerns in UK Court Cases

The first section of the data analysis consists of a systematic review of court cases in the UK where individuals have used free speech rights to contest the refusal of entry into the country, the refusal of naturalisation, and deportation due to the Home Secretary citing extremism concerns.

This section examines how 'extremism' is defined in legal discourse, particularly in terms of public order laws, unacceptable behaviour policies, and judgements exercised around protection of the public good. The examples in this section, combined with existing policies around extremism online, inform the framework to follow.

Between 5 June and 10 July 2019, seven freedom of information (FOI) requests were sent to the Home Office regarding information on how many non-nationals were banned, excluded, or deported from the UK between January 2009 and January 2019 in order to be prevented from speaking on public platforms, on the grounds that them speaking was not conducive to the public good (according to the Home Office's published policy on unacceptable behaviours), as well as non-nationals refused naturalisation due to concerns about their extremist views or behaviour.⁴³

The requests further enquired whether these non-nationals had successfully challenged Home Office decisions on the grounds that the decisions violated their right to freedom of expression under Article 10 of the ECHR.

All of the requests except one were refused. This request did not warrant a response, as information on how many non-nationals have been banned from entering the UK due to their extremist views or behaviour was due to be published in the 2019 Annual Transparency Report on Disruptive and Investigatory Powers, and as such, this information was exempt from FOI requests.

The purpose of the requests was to identify a number of cases that could be analysed for extremism and freedom of speech concerns in the UK. As this could not be achieved using FOIs, an in-depth case study approach was used based on publicly available reporting of cases.

Ten cases were identified on the British and Irish Legal Information Institute (BAIIL) database during this time period, of which five were selected to illustrate individuals who were denied opportunities to address public audiences due to extremist behaviour, or had requests such as naturalisation or residency denied due to their extremist speech.

While the UK government's definition of extremism was used in the online search to locate these cases, transcripts reflecting decisions in both the judgement and the appeals process further informed what does and does not constitute extremism.⁴⁴ Moreover, specific factors used in the decision-making process within the case studies, such as burden of proof, history of the speaker, repeated instances of extremist speech, incitement, and repudiation of extremist views, helped to inform the resulting framework.

⁴³ The unacceptable behaviours policy was outlined by the then Home Secretary to exclude extremist speakers likely to foment terrorism from the country after the 7 July 2005 London bombings. Exclusion could include, but not be limited to, banning entry into the UK, deportation, or refusal of immigration status.

⁴⁴ As jurisprudence built on government definitions.

2.2. Case Study 1: Zakir Naik

Online Influence						
Platform	Active Since	Location	Likes	Followers	Comments	Shares
Facebook	7 June 2013	Kuala Lumpur	17,511,434	17,485,045	Up to 1,800 per post, rarely exceed 350 per post	Up to 5,100 per post, rarely exceed 2,000 per post
Twitter	May 2012 (inactive since December 2018)	Mumbai	10	189,000	Up to 305 per post, rarely exceed 40 per post	Up to 199 per post, rarely exceed 150 per post
YouTube	5 January 2011	Malaysia	Not applicable	1,250,000	Up to 6,834 per video, rarely exceed 700 per video	Up to 564,000 per video, rarely exceed 40,000 per video

2.2.1. Influence

Zakir Naik, from India, is described in court documents as a prominent Muslim writer and public speaker.⁴⁵ In 2010, he was ranked as 89 of 100 most powerful Indians by the Indian Express.⁴⁶ From 2007 to 2011, Naik hosted an annual ‘Peace Conference’ in Mumbai, attracting an audience of over one million to listen to his speeches.⁴⁷ Naik’s YouTube channel, active since 2011, has over 74 million views.

2.2.2. Extremism Concerns

Since 1997, Naik has made a number of controversial and extremist statements, including that “every Muslim should be a terrorist”,⁴⁸ and that every Muslim should support Osama bin Laden’s attacks on the United States (US).⁴⁹ Naik has denounced the US as “the biggest terrorist”,⁵⁰ and stated that Jews are enemies of Islam and that they control the politics of the US.⁵¹ In 2012, Indian police denied Naik permission to hold his annual ‘Peace Conference’ due to security concerns; it has not taken place since.⁵² In 2016, Indian intelligence and counterterrorism

⁴⁵ “Dr Naik, the claimant, is a leading Muslim writer and public speaker” in Justice Cranston, ‘Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India’, *BAILII*, 5 November 2011, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Admin/2010/2825.html&query=\(zakir\)+AND+\(naik\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Admin/2010/2825.html&query=(zakir)+AND+(naik)) last visited: 18 June 2019. See also: “Dr Naik is a Muslim speaker of international reputation” in Lord Justice Carnwath, Lord Justice Jackson and Lord Justice Gross, ‘Between: The Queen on the Application of Naik and Secretary of State for the Home Department’, *BAILII*, 19 December 2011, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2011/1546.html&query=\(zakir\)+AND+\(naik\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2011/1546.html&query=(zakir)+AND+(naik)), last visited: 19 June 2019.

⁴⁶ Justice Cranston, ‘Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India’ *BAILII*, 5 November 2011. See also clarification that the *Sunday Express* is Indian-based in web archives: <http://archive.indianexpress.com/news/the-most-powerful-indians-in-2010-no.-8190/575690/0>.

⁴⁷ Ibid.

⁴⁸ Justice Cranston, ‘Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India’, *BAILII*, 5 November 2011.

⁴⁹ Ibid.

⁵⁰ Ibid.

⁵¹ Ibid.

⁵² Ibid; Goel Sharma, S., ‘Meet Zakir Naik, The Doctor-Turned-Islamic Preacher From Mumbai Who ‘Inspired’ Dhaka Terrorists’, *Scoop Whoop*, 5 July 2016, available at: <https://www.scoopwhoop.com/Meet-Zakir-Naik-The-DoctorTurnedIslamic-Preacher-From-Mumbai-Who-Inspired-Dhaka-Terrorists/>, last visited: 15 August 2019.

agencies began investigating connections between Naik and terrorist organisation Lashkar-e-Taiba (LeT), in relation to his role in inspiring terrorists involved in the 2008 Mumbai terror attacks, whose perpetrators listened to and circulated his speeches.⁵³ Naik heads organisations such as the Islamic Research Foundation (IRF), which was banned in India in 2017 to safeguard national security,⁵⁴ and previously headed broadcasting companies such as Lord Production Inc Ltd.⁵⁵ The stated aim of these organisations is to promote the faith and practice of Islam, but Naik's Peace TV, the broadcast licence of which is owned by Lord Production, is banned in India, Bangladesh, and Canada.⁵⁶

2.2.3. Banned Entry to UK due to Extremism Concerns

Between May 1990 and July 2009, Naik undertook 15 visits to the UK.⁵⁷ However, prior to a series of planned lectures expected to attract an audience of approximately 45,000 people, Naik was investigated by the Office for Security and Counterterrorism (OSCT), which led to a Home Office decision to exclude him from entering the UK on 16 June 2010.⁵⁸ In the letters sent to Naik, the Home Secretary justified the decision on the basis that Naik had engaged in unacceptable behaviour by making statements justifying terrorism and fomenting hatred, particularly his comments on terrorism, Osama bin Laden, and Jews, and was likely to do so again if granted entry. Naik's exclusion was therefore deemed conducive to the public good,⁵⁹ and the decision to exclude him was seen as one that would reduce his ability to publicly address his audience.

An interesting facet of the case emerged in a letter to MPs by the Rt Hon Baroness Neville-Jones, then security minister, on 17 June 2010, explaining the decision in light of representations MPs might receive from constituents.⁶⁰ In this letter, it was stated that Naik would be able to make representations to the Home Office demonstrating he had clearly and consistently repudiated his extremist views for the purposes of challenging the decision to refuse him entry.⁶¹ Naik lodged a judicial review on 12 August 2010, as a result of which his exclusion was deemed lawful.⁶² In statements for the purpose of these proceedings, Naik attempted to dissociate himself from extremist views, emphasising his absolute opposition to terrorism and that, following 9/11, he had not repeated his favourable comments about Osama bin Laden.⁶³

⁵³ Ibid; Kumar, A., Sawant, G.C., 'EXPOSED: Zakir Naik's link to 26/11 mastermind Hafiz Saeed', *India Today*, 7 July 2016, available at: <https://www.indiatoday.in/india/story/exposed-zakir-naiks-link-to-26-11-mastermind-hafiz-saeed-327940-2016-07-07>, last visited: 18 July 2019.

⁵⁴ Khan, S. 'India bans Zakir Naik's Islamic Research Foundation to "safeguard national security"', *The Independent*, 17 March 2017, available at: <https://www.independent.co.uk/news/world/asia/zakir-naik-india-ban-islamic-research-foundation-national-security-safeguard-muslims-a7633566.html>, last visited 1 November 2019.

⁵⁵ Justice Cranston, 'Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India' *BAILII*, 5 November 2011. It is important to note that while the notes from both the High Court and the Court of Appeal references the company in question as "Lords Production Inc Ltd", the Companies House register lists it as "Lord Production Inc Ltd". For more, see: <https://beta.companieshouse.gov.uk/company/05414465>. Furthermore, Naik resigned his directorship of Lord Production Ltd on 1 May 2018.

⁵⁶ Webb, E., 'Wolves in sheep's clothing: How Islamist extremists exploit the UK charitable sector', The Henry Jackson Society (2018), available at: <http://henryjacksonsociety.org/wp-content/uploads/2018/02/HJS-Islamist-Charity-Report.pdf>, last visited: 18 July 2019, pp.46-47.

⁵⁷ Justice Cranston, 'Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India', *BAILII*.

⁵⁸ Ibid.

⁵⁹ Ibid. In a letter in August 2010 responding to Dr Naik's initial objection to this ruling, the Home Secretary further clarified her position by emphasising that while an escalation of community tensions could arise both due to Dr Naik's presence and due to his exclusion, the Home Office had weighed these possibilities and determined that the escalation arising due to his presence would be greater.

⁶⁰ Ibid.

⁶¹ Ibid.

⁶² Ibid.

⁶³ Ibid.

Naik also claimed that his anti-Semitic statements formed part of a wider speech criticising the idea that Muslims and Jews were enemies.⁶⁴ These claims were dismissed by the High Court as marginalising the significance of his statements or using semantics to understate their significance.⁶⁵ In 2011 Naik appealed the decision of the High Court, with the appeal being rejected by the Court of Appeal in October 2011 and his exclusion upheld.⁶⁶

2.2.4. Extremism Definitions and Justifications

This case is crucial for a number of reasons. First, it sheds light on Naik's influence as a speaker, and the actions taken to limit this influence. Second, it assists in understanding the territorial limits of the right of freedom of expression, a point which is especially relevant to technology companies that have to operate in international contexts. Third, the case illustrates an overall lack of cohesion between decisions taken to limit a speaker's speech in the real world, as compared to the online space.

It was observed by the Court of Appeal that Article 10 of the ECHR provides for freedom of expression without interference of public authority, but this may be subject to restrictions in the interests of national security, public safety, prevention of disorder and crime, and protection of health and morals.⁶⁷ In certain cases, including Naik's, freedom of expression could be restricted if this was in the public interest, namely, the right of the public to be protected from either an extremist speaker or the speech resulting from such an individual. Despite Naik's objections that his extremist statements were of little significance, his defence was invalid as, due to his prominence as a speaker, his statements could gain in significance by being taken out of context.⁶⁸ It should also be acknowledged that as well as Naik's prominence as a speaker, his appeal also failed because the Court did not believe his attempts to dissociate himself from extremist views. This has implications for any restrictions on freedom of speech in the online sphere, as it establishes the principle that even when a speaker obscures and qualifies extremist content, this content may still be restricted if a speaker is sufficiently prominent.

Naik's appeal alleged an infringement of his right to freedom of expression under Article 10 of the ECHR. During the High Court proceedings, it was ruled that the Convention applies only within the territory of a relevant member state, meaning Naik's right to freedom of expression would only be protected by the Convention when he was physically present in the UK.⁶⁹ In contrast, at the Court of Appeal this decision was brought into question by the suggestion that Article 10 is not as strictly limited by territoriality as other parts of the Convention.⁷⁰ Therefore, unlike other rights under the Convention, states could be responsible for protecting an individual's freedom of speech, even when that individual was not yet present on their territory. Despite this confusion, no final decision was reached on the territoriality question, on the basis that it was not necessary for the decision on the case.⁷¹ Instead, it was acknowledged by the Court

⁶⁴ Ibid.

⁶⁵ Ibid.

⁶⁶ Lord Justice Carnwath, Lord Justice Jackson and Lord Justice Gross, 'Between: The Queen on the Application of Naik and Secretary of State for the Home Department', *BAI/II*, 19 December 2011, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2011/1546.html&query=\(zakir\)+AND+\(naik\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2011/1546.html&query=(zakir)+AND+(naik)), last visited: 19 June 2019.

⁶⁷ Ibid.

⁶⁸ Ibid. "The very prominence of Dr Naik, as an international figure of great influence, may mean that isolated statements even taken out of context take on a much greater significance. As he himself acknowledges, it is the 'curse' of substantial religious leaders that their words may be 'appropriated by fanatical extremists'."

⁶⁹ Ibid. The ruling was made on the basis of *R(AI-Skeini) v Secretary of State for Defence [2008] 1 AC 153* and *R (Smith) v Oxfordshire Assistant Deputy Coroner [2010] UKSC 29*.

⁷⁰ The ruling was made on the basis of cases such as *Cox v Turkey [2010] Imm AR 4*.

⁷¹ Lord Justice Carnwath, Lord Justice Jackson and Lord Justice Gross, 'Between: The Queen on the Application of Naik and Secretary of State for the Home Department', *BAI/II*, 19 December 2011.

of Appeal that Article 10 rights were enjoyed by Naik's supporters in the UK, including the right to receive information.⁷² As such, regardless of whether the British government was responsible for protecting Naik's freedom of expression, it was responsible for the right of his audience to consume his speeches.⁷³ This carries significant implications for the online space, including Naik's own social media presence, as an individual expressing extremist views online from outside the UK could cite the right of British audiences to access their extremist content in appealing against restrictions to their social media presence.

The decision to exclude Naik was based on the list of unacceptable behaviours outlined by the Home Secretary at the time. This indicative but not exhaustive list envisaged the exclusion of those who foment terrorism or encourage terrorist activity in others, provoke other serious criminal acts or cause inter-communal violence through public speaking, running a website, publishing material or using the position of teacher.⁷⁴

The latter designation was indicative of the influence and authority of the individual at the time of the decision regarding their entry to the UK, but still applies with respect to extremism-related convictions in present day. Mahmudul Choudhury, a teacher from Tower Hamlets, was in 2015 convicted of a racially aggravated offence for an anti-Semitic Facebook post praising Hitler, resulting in a professional life ban.⁷⁵ The resulting sentence was indicative of the individual's position of power and authority, similar to exclusion of extremist speakers.

The judgement suggested a more complex solution than the simple binary of banning or not banning an extremist individual from speaking. Interference with rights under Article 10 was justified on the grounds that it was made in accordance with a predictable legal basis, the list of unacceptable behaviours⁷⁶ and proportionate to the legitimate aims of maintaining community cohesion that the state was pursuing. Although Naik's controversial presence in the UK had been prevented, he could still disseminate his views through formats such as Peace TV, thereby enabling his listeners to receive his information.⁷⁷

Offline, this set a precedent for a graded approach in which the degree of restrictions imposed is adjusted to the level of harm the speech is expected to cause. Such an approach entails subjecting decisions to careful scrutiny by establishing the context of the prescribed expression to determine whether language was intended to incite violence and whether there was a real risk of violence, as well as issues of the author's influence, prominence in the media, and proximity to the centre of violence.⁷⁸

A graded approach also carries implications for restrictions on extremist speech in the online sphere. The influence a speaker has online, possibly measured in popularity or number of

⁷² The Judge commented that Article 10 rights were (without deciding the issue) *possibly* also enjoyed in respect of Naik personally. Justice Cranston, 'Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India', *BAILII*, 5 November 2011.

⁷³ *Ibid.* This was seen by the Court as particularly important as Dr Naik's speeches usually include a significant interactive element in which he answers questions from members of the audience.

⁷⁴ *Ibid.*

⁷⁵ 'Mr Mahmudul Choudhury Professional conduct panel outcome: Panel decision and reasons on behalf of the Secretary of State for Education, October 2015', HM Government (2015), available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/468992/Web_decision_-_Choudhury__Mahmudul_-_0652224.pdf, last visited: 18 September 2019. This case was coded and included in the extremism indicator index.

⁷⁶ Lord Justice Carnwath, Lord Justice Jackson and Lord Justice Gross, 'Between: The Queen on the Application of Naik and Secretary of State for the Home Department', *BAILII*, 19 December 2011.

⁷⁷ Justice Cranston, 'Between: Dr Zakir Naik and (1) The Secretary of State for the Home Department (2) Entry Clearance Officer, Mumbai, India', *BAILII*, 5 November 2011.

⁷⁸ Lord Justice Carnwath, Lord Justice Jackson and Lord Justice Gross, 'Between: The Queen on the Application of Naik and Secretary of State for the Home Department', *BAILII*, 19 December 2011.

followers, can affect the degree to which their speech should be restricted. The context in which extremist language is used is also important, possibly both in relation to events offline such as a moment of particularly high communal tension, and in relation to the way in which the speech is used online in response to these events. It also requires policymakers to consider how proximity to the centre of violence can be judged in the online sphere, where there is less connection between a speaker's location and their audience.

Although Naik has been excluded from entering the UK, this decision has not translated to the online sphere. Rather, his Facebook and Twitter accounts are easily accessible, with the former featuring a video in which Naik describes Jews as, "as a whole", enemies of Islam.⁷⁹ Naik's lectures can also be accessed through YouTube, including material in which he endorses the idea that a man may marry a girl who has reached the age of 13.⁸⁰

Additionally, Naik's Peace TV has kept its UK broadcasting licence, in spite of his exclusion in 2010, and continues to be broadcast to millions of British households.⁸¹ This demonstrates an overall lack of coordination between the media (including social media sites) and the Home Office. Such a vacuum may be filled with the creation of a new framework where multiple bodies work together to identify extremist speakers and content.

2.3. Case Study 2: Pamela Geller and Robert Spencer

Online Influence						
Pamela Geller						
Platform	Active Since	Location	Likes	Followers	Comments	Shares
Facebook	Not known, error on Facebook page	New York	1,327,609	1,321,018	Up to 578 per post, rarely exceed 150 per post	Up to 2,500 per post, rarely exceed 500 per post
Twitter	July 2008	Not known	360	199,000	Up to 346 per post, rarely exceed 40 per post	Up to 551 per post, rarely exceed 400 per post
YouTube	28 April 2006	United States	Not applicable	28,900	Up to 103 per video, only three videos in past month	Up to 5,500 views

⁷⁹ Available at: <https://en-gb.facebook.com/zakirnaik/>, last visited: 18 July 2019; available at: <https://twitter.com/zakirnaik?lang=en>, last visited: 18 July 2019; video available at: <https://twitter.com/zakirnaik/status/792237338382307328>, last visited: 18 July 2019.

⁸⁰ Available at: <https://twitter.com/zakirnaik/status/792237338382307328>, last visited: 18 July 2019.

⁸¹ Baynes, C., "Hate preacher" banned from UK still broadcasting to millions in Britain through "personal TV station", *The Independent*, 21 November 2018, available at: <https://www.independent.co.uk/news/media/hate-preacher-banned-uk-zakir-naik-ofcom-islamic-extremism-peace-tv-a8644931.html>, last visited: 18 July 2019.

Robert Spencer						
Platform	Active Since	Location	Likes	Followers	Comments	Shares
Facebook	25 July 2014	Not known	21,227	22,272	Up to 83 per post, rarely exceed 50 per post	Up to 391 per post, rarely exceed 100 per post
Twitter	February 2009	Not known	284	126,000	Up to 129 per post, rarely exceed 50 per post	Up to 698 per post, rarely exceed 200 per post
YouTube as Jihad Watch Video	4 October 2008 (inactive since November 2018)	Not known	Not applicable	27,600	Up to 1,345 per video, rarely exceeded 250 per video	Up to 33,000 per video, rarely exceeded 15,000 per video

2.3.1. Influence

Pamela Geller and Robert Spencer, co-founders of Freedom Defense Initiative,⁸² an organisation accused of Islamophobia,⁸³ have publicly stated their opposition to the spread of extreme Islam and *sharia* law in Western societies.⁸⁴ Neither has any convictions in the UK or US.⁸⁵ Geller provoked controversy after funding an advertising campaign in San Francisco and New York that was interpreted to imply that Muslims are “savages”, as well as for displaying an offensive image of Prophet Mohammed online.⁸⁶ Spencer has argued that all schools of Islam teach warfare against non-Muslims, and that Islam is inherently violent.⁸⁷ For these reasons, both speakers are considered Islamophobic.⁸⁸ Geller’s YouTube channel, active since 2006, has over nine million views. Spencer’s YouTube channel, operating under JihadWatchVideo, has been active since 2008 and has over 3.4 million views. His videos are also shared through a number of fan accounts.

2.3.2. Banned Entry to UK due to Extremism Concerns

In 2013, Geller and Spencer planned to attend a rally by the English Defence League (EDL) in Greenwich to coincide with the Armed Forces Day celebration at Woolwich Barracks.⁸⁹ The

⁸² Also known as Stop the Islamization of America. Lord Justice Patten, Lord Justice Tomlinson and Lord Justice Floyd, ‘Between the Queen on the application of Pamela Geller and Robert Spencer and Secretary of State for the Home Department’ *BAILII*, 5 February 2015, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2015/45.html&query=\(pamela\)+AND+\(geller\)+AND+\(robert\)+AND+\(spencer\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2015/45.html&query=(pamela)+AND+(geller)+AND+(robert)+AND+(spencer)), last visited: 19 June 2019.

⁸³ The accusation was made by the Southern Poverty Law Center; Miller, M.E., ‘Killing Jews is Worship’ posters will soon appear on NYC subways and buses’, *The Washington Post*, 22 April 2015, available at: <https://www.washingtonpost.com/news/morning-mix/wp/2015/04/22/killing-jews-is-worship-posters-will-soon-appear-on-nyc-subways-and-buses/>, last visited: 8 August 2019. See also, Lord Justice Patten, Lord Justice Tomlinson and Lord Justice Floyd, ‘Between the Queen on the application of Pamela Geller and Robert Spencer and Secretary of State for the Home Department’ *BAILII*, 5 February 2015, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2015/45.html&query=\(pamela\)+AND+\(geller\)+AND+\(robert\)+AND+\(spencer\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2015/45.html&query=(pamela)+AND+(geller)+AND+(robert)+AND+(spencer)), last visited: 19 June 2019.

⁸⁴ Lord Justice Patten, Lord Justice Tomlinson and Lord Justice Floyd, ‘Between the Queen on the application of Pamela Geller and Robert Spencer and Secretary of State for the Home Department’ *BAILII*.

⁸⁵ *Ibid.*

⁸⁶ *Ibid.*

⁸⁷ *Ibid.*

⁸⁸ *Ibid.*

⁸⁹ *Ibid.* Separately, it is important to note that the EDL as an organisation has been banned from the Facebook platform. For more, see: Hern, A., ‘Facebook bans far-right groups including BNP, EDL and Britain First’, *The Guardian*, 18 April 2019, available at: <https://www.theguardian.com/technology/2019/apr/18/facebook-bans-far-right-groups-including-bnp-edl-and-britain-first>, last visited: 11 December 2019.

EDL is a far-right organisation whose stated aim is to oppose Islamism and which has been involved in violent clashes with both anti-EDL groups such as Unite Against Fascism (UAF) and the police.⁹⁰ As a result, on 24 June 2013, the Metropolitan Police wrote to the Home Secretary asking for the exclusion of Geller and Spencer from entry into the UK.⁹¹ The letter noted that the timing of the rally was particularly sensitive in light of the murder of soldier Lee Rigby by Islamist terrorists the same year.⁹² The letter argued that if Geller and Spencer addressed the rally, their views on Islam and Geller's outspoken support of Israel could attract Muslim and pro-Palestinian groups.⁹³ This would complicate policing operations, which envisaged the need to keep the EDL and UAF apart.⁹⁴ As attendance could undermine community cohesion and foster violence, Geller and Spencer's presence was seen as not conducive to the public good.⁹⁵

On 25 June 2013, the Home Secretary wrote to Geller and Spencer informing them that they were banned from entering the country on the basis of the unacceptable behaviours policy. The letter from the Home Office noted statements made by Geller and Spencer that equated Islam as a whole with fighting *jihad*, and described Freedom Defense Initiative as a hate group.⁹⁶ It therefore concluded that their presence was not conducive to the public good.⁹⁷ On 3 July, Geller and Spencer gave notice of their intention to bring judicial review proceedings.⁹⁸ The hearing, on 20 November 2014, was based on the grounds that they did not intend to incite violence, and that their rights under the Convention had been violated.⁹⁹ The appeal was dismissed, and the decision of the Upper-tier Tribunal (UtT) to refuse permission to apply for judicial review was upheld.¹⁰⁰

2.3.3. Extremism Definition and Justifications

This case was important for three reasons. First, and relevant to the work of regulating such speakers and organisations online, was the Home Secretary's relative discretion in making decisions on limiting freedom of expression. The Court of Appeal ruled that while decisions on freedom of expression should be subject to careful scrutiny, the Home Secretary's decision carried special weight.¹⁰¹ As with *Naik*, such decisions were made with access to information and advice that the Court – as well as ordinary members of the public – did not have.¹⁰² Also similar to *Naik* was the justification that the decision to interfere with rights under Article 10 was deemed proportionate to a potential risk of violence, and as such, the Court would hesitate before substituting its own assessment of the threat.¹⁰³ This is relevant to regulating such speakers online, as technology companies would need to work in close conjunction with

⁹⁰ Ibid.

⁹¹ Ibid.

⁹² Ibid.

⁹³ Ibid.

⁹⁴ Ibid.

⁹⁵ Ibid.

⁹⁶ Ibid.

⁹⁷ Ibid.

⁹⁸ Ibid.

⁹⁹ Ibid.

¹⁰⁰ Ibid.

¹⁰¹ This decision was made on the basis of *R (on the application of Lord Carlile of Berriew QC and others) v Secretary of State for the Home Department*, [2014] UKSC 60.

¹⁰² Lord Justice Patten, Lord Justice Tomlinson and Lord Justice Floyd, 'Between the Queen on the application of Pamela Geller and Robert Spencer and Secretary of State for the Home Department', *BAILII*, 5 February 2015. In keeping with the democratic values of the Convention, decisions on public order in which the costs of failure are high should be made by elected officials, rather than judges by virtue of being democratically accountable for their choices.

¹⁰³ Ibid.

the police and courts to obtain information about which speakers are problematic, as much of this material would not be publicly available for security reasons.¹⁰⁴ Moreover, importance was given to incitement of violence or the risk of violence inherent in the words expressed by an influential extremist speaker, something that social media companies would have to be especially cautious about regulating in the online space.

Second, the case illustrated that it is primarily the responsibility of the state (acting through the police) to prevent violence that may arise from a speaker's words.¹⁰⁵ This was an issue that had not arisen in *Naik* due to the speaker's lack of immediate association with an organisation involved in violent incidents similar to the EDL. Although the issue of the state's responsibility to prevent violence resulting from a speaker's words was discussed by the Court of Appeal primarily in relation to immigration rather than freedom of expression laws, it was also cited in relation to justifying restrictions on freedom of expression.¹⁰⁶

An immediate association with violence relied on the premise that a speaker's words can be as dangerous as the action itself, whether or not the action actually results from these words. This is particularly relevant to social media companies, who may need to consider regulating speech for its potential to inspire or cause violence offline.¹⁰⁷ Such immediacy may be difficult to measure or map in the online space, as compared to the offline space where immediate unlawful violence could be measured in terms of scheduled rallies and demonstrations.¹⁰⁸

Finally, the case established the practical status of the unacceptable behaviours policy, on the basis of which restrictions on freedom of expression are made. This was an issue that had not been discussed in, but could also apply to, *Naik*. A major aspect of Geller and Spencer's appeal was that the policy was not lawful because it had not been laid before Parliament, and the Home Secretary had to obtain approval from Parliament for any changes to rules regulating entry into the UK.¹⁰⁹

However, the Court noted a difference between new rules and the application of existing rules,¹¹⁰ and decided that as the policy only constituted a set of guidelines, it did not set new conditions that needed to be satisfied for an individual to enter the UK.¹¹¹ This meant that the unacceptable behaviours policy was not a new law regulating entry into the UK, but rather helped to implement existing laws. As such, a definition of extremism could also be interpreted and exercised upon discretion of the Home Secretary.

¹⁰⁴ It should also be acknowledged that such restrictions and limitations apply to research that is based on court judgements as well (such as this report), given that it may be difficult to extrapolate key factors from publicly available judgements, which may not contain all required information to understand why an individual is deemed an extremist.

¹⁰⁵ Restrictions on the right to free speech under the US Constitution First Amendment can also be justified, but the threshold is higher (see *Brandenburg v Ohio*).

¹⁰⁶ Lord Justice Patten, Lord Justice Tomlinson and Lord Justice Floyd, 'Between the Queen on the application of Pamela Geller and Robert Spencer and Secretary of State for the Home Department', *BAILII*, 5 February 2015.

¹⁰⁷ It should be noted that some technology companies already have policies against credible threats and incitement. See, for example: 'Dangerous individuals and organisation', *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/dangerous_individuals_organizations/, last visited: 17 December 2019; 'Violence and incitement', *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/credible_violence, last visited: 18 December 2019; and 'Proposing or publicising crime', *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/promoting_publicizing_crime, last visited 17 December 2019.

¹⁰⁸ Despite this, platforms such as Facebook can still work closely with police to understand how advertising for an offline event such as a rally or demonstration may be manifesting itself online, through an event page or group. Moreover, the indicators "Mentions resorting to violence with personal agency", and "Incites violence", and "Incites hatred", as well as repeated instances of sharing such views with intent online, may assist in understanding the immediacy of unlawful violence in the online realm.

¹⁰⁹ This is demanded by section 3(2) of the Immigration Act 1971.

¹¹⁰ This decision was made on the basis of *R (Alvi) v Secretary of State for the Home Department*, [2012] 1 WLR 2208.

¹¹¹ Lord Justice Patten, Lord Justice Tomlinson and Lord Justice Floyd, 'Between the Queen on the application of Pamela Geller and Robert Spencer and Secretary of State for the Home Department' *BAILII*, 5 February 2015.

2.4. Case Study 3: Raed Salah Mahajna

Online Influence						
Raed Salah Mahajna						
Platform	Active Since	Location	Likes	Followers	Comments	Shares
Facebook	7 September 2014 (inactive since October 2016)	Not known	57,401	57,482	Up to 85 per post, rarely exceeded 40 per post	Up to 90 per post, rarely exceeded 50 per post
Twitter	Not available	Not available	Not available	Not available	Not available	Not available
YouTube	Not available	Not available	Not available	Not available	Not available	Not available

2.4.1. Influence

Raed Salah Mahajna, commonly known as Raed Salah, is a Palestinian and an Israeli national who leads the northern branch of the Islamic Movement in Israel (an Islamist movement), has been mayor of the Israeli town Umm al-Fahm, and has made multiple visits to the UK since 1990 in order to speak at conferences.¹¹² He has previously travelled in Europe without causing disorder.¹¹³

2.4.2. Extremism Concerns

In 2011, Salah was invited to the UK by Middle East Monitor (MEMO) to deliver a speech at the House of Lords.¹¹⁴ In June of that year, it was decided by the Home Secretary that Salah's presence would not be conducive to the public good, and he would therefore be denied entry.¹¹⁵ This decision was taken on the basis of Salah's historical statements, which could lead to inter-communal violence in Britain and were therefore proscribed under the unacceptable behaviours list, which after 2011 became part of the Prevent strategy.¹¹⁶ These statements included an anti-Semitic poem published in 2003, invocation of the blood libel in 2007, and the promotion of martyrdom in 2009.¹¹⁷ It was also noted that the Islamic Movement in Israel has connections to Hamas;¹¹⁸ the military wing of Hamas was proscribed as a terrorist organisation in the UK in March 2001.

However, neither Salah nor immigration officers were made aware of the decision to deny him entry, and Salah was able to enter the UK on 25 June 2011.¹¹⁹

On 28 June 2011, Salah was arrested for the purposes of being deported.¹²⁰ On 29 June 2011, he appealed to the First-tier Tribunal (FtT) against the decision of the Home Secretary to make a

¹¹² Ockelton, C.M.G. and Judge Pitt, 'Between Raed Salah Mahajna and the Secretary of State for the Home Department', *BAILII*, 16 April 2012, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/uk/cases/UKUT/IAC/2012/B1.html&query=\(raed\)+AND+\(salah\)+AND+\(mahajna\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/uk/cases/UKUT/IAC/2012/B1.html&query=(raed)+AND+(salah)+AND+(mahajna)), last visited: 20 June 2019.

¹¹³ *Ibid.*

¹¹⁴ *Ibid.*

¹¹⁵ *Ibid.*

¹¹⁶ *Ibid.*

¹¹⁷ Nicol, The Hon. Mr. Justice., 'Between Raed Salah Mahajna and the Secretary of State for the Home Department', *BAILII*, 30 September 2011, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Admin/2011/2481.html&query=\(mahajna\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Admin/2011/2481.html&query=(mahajna)), last visited: 5 November 2019.

¹¹⁸ *Ibid.*

¹¹⁹ *Ibid.*

¹²⁰ Ockelton, C.M.G. and Judge Pitt, 'Between Raed Salah Mahajna and the Secretary of State for the Home Department', *BAILII*, 16 April 2012.

deportation order against him.¹²¹ The FtT dismissed his appeal.¹²² During proceedings, it was observed that Salah's poem had been incorrectly translated, and was deemed to be about oppression, rather than being anti-Semitic. The Tribunal also observed that, in the speech of 2007, Salah did not make the blood libel, and that there was no clear evidence of Salah's speech on martyrdom provoking violence. In spite of these qualifications, the Tribunal dismissed his objections, as the infringement on Salah's rights under Article 10 of the ECHR was deemed proportionate by the FtT, especially as it would not prevent Salah from communicating through modern communications methods.¹²³ Salah made an appeal against the exclusion order, for which a hearing was held on 6 and 8 February 2012. The deportation order was overturned.¹²⁴ The case demonstrates the principle that courts grant the Home Secretary less discretion on curtailing freedom of expression in appeals such as this one, than in judicial reviews such as those in the previous cases of *Naik* or *Geller and Spencer*.

2.4.3. Extremism Definition and Justifications

What was particularly interesting in the Salah case was the appeals process. The Upper-tier Tribunal (UtT) observed that while it was responsible for reviewing the proportionality of any restriction on an individual's rights under Article 10, it would not usually contradict the Home Secretary's assessments of what is conducive to the public good.¹²⁵ However, it observed that courts should still weigh these decisions against competing considerations, rather than deferring to them.¹²⁶ Unlike a judicial review as in *Naik*, the Tribunal was not confined to material available to the Home Secretary at the time of their decision, but also considered material arising after that date.¹²⁷ In light of this, the Tribunal observed that new, more accurate information about Salah's statements could not support the Home Secretary's earlier decision.¹²⁸

It was ruled that unlike *Naik*, Salah did not have a clear and potentially dangerous agenda in his public pronouncements. Instead, the statements on the blood libel and martyrdom were not deemed to be at the core of his message, but rather isolated statements on a single occasion. It was noted that as his presence in the UK had never prompted hatred or intercommunal violence, it was less necessary for the Home Secretary to predict what the results of his presence would be.¹²⁹ As such, Salah's case is different from precedents set in *Naik* and *Geller and Spencer*.

In both of those cases, the decision to exclude extremist individuals was not made on the basis that they had no history of previously causing violence, but on the assumption that they may do so in the future. This reflects an inconsistency in how the principle of treating an appellant's

¹²¹ Ibid.

¹²² Judge Renton and Judge C Lloyd, 'Between Raed Salah Mahajna and the Secretary of State for the Home Department', *BAILII*, 26, 27 September and 3 October 2011, available at: [https://www.bailii.org/uk/cases/UKUT/IAC/2012/B1\(image1\).pdf](https://www.bailii.org/uk/cases/UKUT/IAC/2012/B1(image1).pdf), last visited 30 October 2019.

¹²³ Ibid.

¹²⁴ Ibid.

¹²⁵ This was decided on the basis of *SSHD v Rehman [2001] UKHL 47*. As with the *Geller and Spencer* case, it was deemed that such decisions by the Home Secretary are made on the basis of better information and advice than that available to the courts. Therefore, it is only necessary for the FtT to determine whether material on the basis of which the Home Secretary could reasonably expect behaviour harmful to the public good exists. For more, see: Ockelton, C.M.G. and Judge Pitt, 'Between Raed Salah Mahajna and the Secretary of State for the Home Department', *BAILII*, 16 April 2012.

¹²⁶ Ockelton, C.M.G. and Judge Pitt, 'Between Raed Salah Mahajna and the Secretary of State for the Home Department', *BAILII*, 16 April 2012.

¹²⁷ Ibid. Further, it was noted that, unlike judicial reviews which are simply a supervision of executive decision-making, courts deal with appeals because they have been given the jurisdiction to do so by legislators and therefore have the constitutional power to remake a decision if necessary.

¹²⁸ Ibid.

¹²⁹ Ibid.

profile as a whole is applied, which would make it more difficult to derive a clear precedent for future online and offline cases on the basis of these rulings alone.¹³⁰

The case is significant for three further reasons. The first is whether the Home Secretary takes into account criminal indictments for actions that have not been proven. In relation to a June 2011 indictment alleging that Salah had incited violence with a sermon in 2007, it was ruled that had the Israeli state demonstrated serious concern about Salah’s words in 2007 then the indictment would have been relevant regardless of its unproven status.¹³¹ However, in light of many years between the sermon and the indictment, and the lack of any similar incidents, the indictment was deemed unworthy of consideration.¹³² Second, the case outlines the degree to which interpretations of a speaker’s statements justify restrictions on their freedom of expression. In ruling that Salah’s exclusion was conducive to the public good, the FtT observed that in order to fall under the list of unacceptable behaviours, a statement need not be racist as such, but only able to foment hatred that could lead to intercommunal violence, either through being virulently critical of a large section of a group (such as the Jewish people), or through being easily interpreted as racist, as in the case of the poem. However, the UtT ruled this was not the case where the interpretation was clearly and wholly inaccurate. As such, the accurately interpreted and reported content of a statement (in this case the poem) had never incited violence.¹³³

Finally, the ruling is significant in relation to the question of exclusion on the basis of ties to an organisation proscribed in the UK. In 2005, Salah was convicted in Israel of ties to Hamas, however, this was deemed insufficient to bring him within the unacceptable behaviours policy, because it was impossible to prove, in both the case in which he was convicted and when assessing his case, that Salah had funded the illegal military, rather than the legal political, wing of Hamas. The facts that the incident occurred long ago and resulted in a brief sentence in Israel were taken into account to suggest that Salah was not considered a danger.¹³⁴ Like the courts, social media companies may face similar difficulties in tracking such individuals, who may have links to extremist groups that have yet to be banned by the Home Secretary, and who take great care in ensuring that such links are hidden from public knowledge. Moreover, varying standards and thresholds in different jurisdictions would pose a further challenge to internationally operating social media platforms.

2.5. Case Study 4: AS and FM

Online Influence						
AS and FM						
Platform	Active Since	Location	Likes	Followers	Comments	Shares
N/A as cases anonymised						

¹³⁰ The Tribunal distinguished *Naik* by stating that the case was distinguished on the basis that “the individual in question had a clear agenda in his public pronouncements that was pervasive and potentially offensive or dangerous” and, unlike in *Naik*, Salah had been in the UK for a prolonged period and it was “less necessary to try to predict what will happen where there is instead evidence of what has (or rather has not) happened”. Such grounds for distinction are based on the assumption that an individual will have a history of making public pronouncements that are extremist. Moreover, while *Salah* was based on the assumption that a history of non-violence is likely to lead to no incitement of violence in the future, the same was not applied to *Geller and Spencer*, and *Naik*, who had also visited the UK previously with no risk of immediate violence. These are therefore not suitable grounds for distinction.

¹³¹ On 23 June 2011, two separate indictments were issued against the Appellant to be eventually heard at the Jerusalem Magistrates’ Court. The allegations contained in the indictments were disputed by the Appellant and were yet to be decided at the time of the FtT proceedings.

¹³² Ibid.

¹³³ Ibid.

¹³⁴ Ibid.

2.5.1. Influence

AS is a Libyan refugee with indefinite leave to remain in the UK, who applied for and had his naturalisation approved in 2008. FM was an Imam who preached at the Hatherley Street Mosque in Liverpool, and was married to a British citizen with whom he had two children.

2.5.2. Extremism Concerns

An initial decision to grant naturalisation to AS was soon after reversed, on the grounds that he had made “extremist Islamist” statements, and therefore did not meet the necessary ‘good character’ requirement for citizenship.¹³⁵ FM applied for and was refused naturalisation on the grounds that he preached anti-Western views, and openly exhibited sympathy for Osama bin Laden, therefore not meeting the necessary ‘good character’ requirements for citizenship.¹³⁶ Along with two other claimants, AHK and AM, the group of four applied for a judicial review of their cases on the grounds that the Home Secretary had provided few or no reasons for not being satisfied that they met the ‘good character’ requirements. The Home Secretary refused to disclose these reasons, as to do so would involve making public information that, if widely known, would be harmful to national security. A hearing was held in January 2012, where the High Court ruled that it could not fairly review the Home Secretary’s decision because it would not have access to the relevant classified material.¹³⁷

All four claimants raised further submissions for judicial review. This led to a second hearing on 22 April 2013, in which AS and FM invoked the Court’s obligation to protect their right to freedom of expression under the ECHR by making them aware of the evidence on the basis of which the Home Secretary had restricted these rights by rejecting their naturalisation. In practice, this meant they were calling for the disclosure of the relevant classified material.¹³⁸ AS’s counsel argued that his refusal of naturalisation had led to him being viewed with considerable suspicion among the close-knit Libyan community. It also led him to be treated more harshly than naturalised Libyans such as the rest of his family, and meant that he was more likely to be searched at airports. As the refusal had been made on grounds of expressing extremist views, AS invoked Article 10 of the ECHR.¹³⁹ FM argued that his refusal of naturalisation affected his reputation. FM’s counsel further argued that the Home Secretary had not stated that FM’s preaching incited violence, and that FM had only “peacefully preach[ed] extreme Muslim views”.¹⁴⁰ He had also not preached to the public at large, but only to a small congregation whose members could attend a different mosque if they preferred. FM therefore invoked Article 9, which protects his right to freedom of religion, and Article 10.¹⁴¹

2.5.3. Extremism Definition and Justifications

This case is significant as it provides insight into whether the defence of rights under Article 10 requires the Home Secretary to disclose matters of national security to justify decisions. The Court observed that, insofar as it involved interference with an individual’s right to freedom of expression in the public interest, this case would follow a similar legal framework to *Naik*. Specifically, the Court noted that the Home Secretary had interfered with FM’s and AS’s rights under Articles 9 and 10. This was because they had been denied the benefit of UK citizenship

¹³⁵ Justice Ouseley, ‘Between AHK, AM, AS, FM and Secretary of State for the Home Department’, *BAILII*, 7 June 2013, available at: [https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Admin/2013/1426.html&query=\(FM\)+AND+\(AS\)+AND+\(2013\)+AND+\(ewhc\)+AND+\(1426\)](https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Admin/2013/1426.html&query=(FM)+AND+(AS)+AND+(2013)+AND+(ewhc)+AND+(1426)), last visited: 20 June 2019.

¹³⁶ *Ibid.*

¹³⁷ *Ibid.*

¹³⁸ *Ibid.*

¹³⁹ *Ibid.*

¹⁴⁰ *Ibid.*

¹⁴¹ *Ibid.*

due to what they had said.¹⁴² The Court noted that there must always be legal protection against arbitrary interference with an individual’s fundamental rights under the ECHR, which meant that decisions to restrict these rights had to be subjected to fair legal review, even in cases where the disclosure of evidence could threaten national security.¹⁴³ However, not disclosing classified evidence was deemed to be consistent even with Article 6 of the ECHR, which guarantees the right to a fair trial.¹⁴⁴ The High Court noted freedoms of religion and expression could be restricted to a greater degree than the right to a fair trial, without this being seen as fundamentally depriving an individual of their rights. This was especially the case given that the Home Secretary was imposing very modest restrictions on the rights to freedom of religion and expression of FM and AS. This referred to the fact that she had only denied them the benefit of naturalisation, rather than restricting their ability to speak or practise their religion outright.¹⁴⁵

2.6. Case Study 5: Alison Chabloz

Online Influence						
Alison Chabloz						
Platform	Active Since	Location	Likes	Followers	Comments	Shares
Facebook	7 March 2014 (currently banned for 12 months)	Not known	801	846	Up to 25 per post, rarely exceeded 10 per post	Up to 29 per post, rarely exceeded 10 per post
Twitter	Not available	Not available	Not available	Not available	Not available	Not available
YouTube	Ordered to abstain from posting on social for 12 months in May 2018, breached order by publishing material to her blog in September 2019 and sentenced to eight weeks’ imprisonment.					

Case 5, that of Alison Chabloz, is unique in that it involves an appeal against restrictions to social media presence under Article 10.

2.6.1. Influence

Alison Chabloz, a prominent British blogger who defines herself as a “Holocaust revolutionist”,¹⁴⁶ has written and performed three songs denying the Holocaust, as well as mocking it by referring to Auschwitz as a “theme park”.¹⁴⁷ The content has been posted online, where she also mocked Holocaust survivors as liars, and attacked the state of Israel as having “no right to exist”.¹⁴⁸ The songs rely on anti-Semitic tropes such as the use of triple parentheses around the title of the song *Survivors*.¹⁴⁹ Two of these songs were first performed at an event held by

¹⁴² Ibid.

¹⁴³ Ibid.

¹⁴⁴ This was decided on the basis of *Carnduff v UK (application 18905/02)*; ‘European Convention on Human Rights’, Council of Europe (1950), p.3.

¹⁴⁵ Justice Ouseley, ‘Between AHK, AM, AS, FM and Secretary of State for the Home Department’, *BAILII*, 7 June 2013.

¹⁴⁶ Bliss, L., ‘Social Media: “A Theme Park just for Fools”, *R v Alison Chabloz* (unreported) Westminster Magistrates’ Court 25 May 2018’, *The Journal of Criminal Law* 82.4 (2018): pp.301-304.

¹⁴⁷ Ibid.

¹⁴⁸ ‘Alison Chabloz has anti-Semitic songs conviction upheld’, *BBC News*, 13 February 2019, available at: <https://www.bbc.co.uk/news/uk-england-derbyshire-47230443>, last visited: 16 July 2019.

¹⁴⁹ Bliss, L., ‘Social Media: “A Theme Park just for Fools”, *R v Alison Chabloz* (unreported) Westminster Magistrates’ Court 25 May 2018’, p.303.

the far-right organisation London Forum in 2016 before being published on Chabloz's blog, while the third was uploaded on YouTube.¹⁵⁰ Her songs were also uploaded by other far-right and Neo-Nazi accounts.

2.6.2. Extremism Definition and Justifications

In May 2018, Chabloz was prosecuted for sending grossly offensive, indecent, obscene or menacing material through a public electronic communications network¹⁵¹ and was convicted of three counts of doing so. She was sentenced to 20 weeks' imprisonment and banned from social media for 12 months. This marked the first UK conviction over Holocaust denial on social media.¹⁵² Chabloz appealed the decision on the grounds that there should be no criminal penalty for singing in "polemical terms" and that, while offensive, her lyrics could not be classified as "grossly offensive". In February 2019, the conviction was upheld by the Southwark Crown Court on the grounds that Chabloz was anti-Semitic and a Holocaust denier.¹⁵³

This case is similar to *Naik* and *Geller and Spencer* as it centres on the use of Article 10 to invoke the right to free speech. At Westminster Magistrates' Court, Chabloz invoked Article 10 of ECHR to argue that her right to freedom of speech was being infringed upon. This defence is identical to the invocation of Article 10 in previously examined cases. The judge rejected this claim on the basis that Chabloz's songs were not an academic critique of the Holocaust or political in nature, but rather an attack on Jewish people.¹⁵⁴ Chabloz sought to defend her speech on the basis that the songs in question were "satire". However, this was refuted by the prosecution which argued that, in their tone and content, the songs were written to cause maximum anti-Semitic offence.¹⁵⁵ This is similar to *Naik*, in which attempts to marginalise the significance of statements were similarly rejected on the grounds that they intended to carry offensive content. This differs from *Salah*, however, where the same defence was successful in overturning free speech restrictions, due to the lack of an established history of causing offence.

The case is further similar to *Naik* and *Geller and Spencer* in that it acts as an aid in determining which forms of speech are not protected by Article 10. However, unlike the other two cases, the ruling restricting Chabloz's rights was framed in terms of hate speech rather than extremism. The Convention, which in Article 10 cases has traditionally given substantial weight to an individual's right to freedom of expression as an important aspect of democratic society, has ruled that Article 10 does not necessarily cover Holocaust denial.¹⁵⁶ The Holocaust is classed as an established historical fact, whose denial is removed from the protection of Article 10 by Article 17, which sets the parameters for how far rights enshrined in the Convention may be restricted.¹⁵⁷ Although there is no specific law against Holocaust denial, it is not protected from restrictions against hate speech online that criminalise intentionally "grossly offensive" material.¹⁵⁸ As such, while Chabloz's speech could be protected under Article 10, the infringement

¹⁵⁰ Ibid.; Poulter, J., 'The Neo-Nazi Home of the UK Alt-Right', *Vice*, 12 March 2018, available at: https://www.vice.com/en_uk/article/9kzyav/the-neo-nazi-home-of-the-uk-alt-right, last visited: 17 July 2019; Chabloz, A., "'Tell Me More Lies" at The London Forum', *Alison Chabloz Blog*, 28 September 2016, available at: <https://alisonchabloz.com/2016/09/>, last visited: 15 August 2019.

¹⁵¹ Under Section 127(1) of the Communications Act 2003 (CA 2003).

¹⁵² 'Alison Chabloz has anti-Semitic songs conviction upheld', *BBC News*, 13 February 2019; Bliss, L., 'Social Media: "A Theme Park just for Fools" R v Alison Chabloz (unreported) Westminster Magistrates' Court 25 May 2018', p.302.

¹⁵³ 'Alison Chabloz has anti-Semitic songs conviction upheld', *BBC News*, 13 February 2019.

¹⁵⁴ 'Alison Chabloz has anti-Semitic songs conviction upheld', *BBC News*, 13 February 2019; Bliss, L., 'Social Media: "A Theme Park just for Fools" R v Alison Chabloz (unreported) Westminster Magistrates' Court 25 May 2018', p.301.

¹⁵⁵ Ibid, p.303.

¹⁵⁶ As demonstrated in *Handyside v United Kingdom* [1976] 1 EHRR 737.

¹⁵⁷ As ruled in *Lehideux and Isorni v France* [1998] ECHR 90; 'European Convention on Human Rights', Council of Europe (1950) p.5; Bliss, L., 'Social Media: "A Theme Park just for Fools" R v Alison Chabloz (unreported) Westminster Magistrates' Court 25 May 2018' pp.302-303.

¹⁵⁸ Under CA 2003; Bliss, L., 'Social Media: "A Theme Park just for Fools" R v Alison Chabloz (unreported) Westminster Magistrates' Court 25 May 2018' p.303.

of Article 10 could be justified. However, the legal justification for restricting Chabloz's speech is different from previously examined cases, as these cases referred to extremist speech or individuals likely to provoke violence listed in the unacceptable behaviours policy, whereas *Chabloz* is characterised as inciting hate, without any reference to this leading to potential violence.

Chabloz lacked a clear legal framework for restricting offensive hate speech, and how such restrictions may differ from those on extremist speech. In spite of the existence of CPS guidelines from 2012 which cite cases intended to illustrate what constitutes the undefined concept of 'grossly offensive' material, confusion continues to exist regarding when an individual goes beyond their right to freedom of expression and engages in criminal activity. Chabloz was therefore not prosecuted for her activity, in spite of numerous reports to the police, until after the charity Campaign Against Antisemitism (CAA) brought forward a private prosecution against her. This example illustrates the larger definitional vacuum that law enforcement and technology companies face when identifying an 'extremist', given that such individuals will employ both hate speech and extremist views to communicate to their audiences. In spite of the emphasis on hate speech rather than extremism in the Chabloz case, the ruling restricted Chabloz's rights under Article 10.

Moreover, the online element of Chabloz's hate speech has meant that the implementation of the ruling against her has been much less effective. Although material directly published by Chabloz cannot be accessed through a simple search on YouTube, a far-right group has republished her song *Survivors* which is easily accessible within the first 10 Google search results.¹⁵⁹ Various videos of Chabloz published by supporters of the far-right, including one in French in which she discusses topics such as Holocaust "revisionism" and anti-Semitic conspiracy theories, are also available on YouTube.¹⁶⁰ In September 2019, Chabloz was jailed for breaching the order by publishing material on her blog, contrary to the advice of probation officers.¹⁶¹ This demonstrates that current restrictions on online content are only partially effective, and continue to allow for the presence of banned speakers.

¹⁵⁹ Google search result on 'Alison Chabloz', available at: https://www.google.com/search?q=alison+chabloz&source=lnms&tbm=vid&sa=X&ved=0ahUKewiEqOXEpbzjAhUVRhUIHZobAPwQ_AUIEigC&biw=1536&bih=754, last visited: 19 November 2019.

¹⁶⁰ YouTube search result on 'Alison Chabloz', available at: https://www.youtube.com/results?search_query=alison+chabloz, last visited: 19 November 2019.

¹⁶¹ Welch, B. 'Holocaust denier Alison Chabloz jailed for continuing to blog despite social media ban', *The JC*, available at: <https://www.thejc.com/news/uk-news/holocaust-denier-alison-chabloz-jailed-for-continuing-to-blog-despite-social-media-ban-1.489185>, last visited: 11 October 2019.

3. Extremism Indicators: A Framework

The following section illustrates major themes to emerge from the five qualitative case studies of individuals denied opportunities to address audiences in the UK due to extremism concerns. Coupled with an examination of 'extremism' and its overlaps with terrorism and hate crime, this section maps various indicators to assess extremist behaviour and speakers online. As such, social media companies can benefit from decisions made by case law in the offline space.

Following the qualitative review of five case studies of extremist speakers, it is apparent that courts and social media companies employ different frameworks to assess extremism.¹⁶² Nonetheless, analysis of these cases points to some important trends in assessing whether, and how, to stop an extremist speaker from addressing an audience. These themes were collectively qualified into the following 20 'extremism indicators', some of which are relevant to activity online, and others which are applicable to behaviour offline.

Figure 3: Indicator Checklist

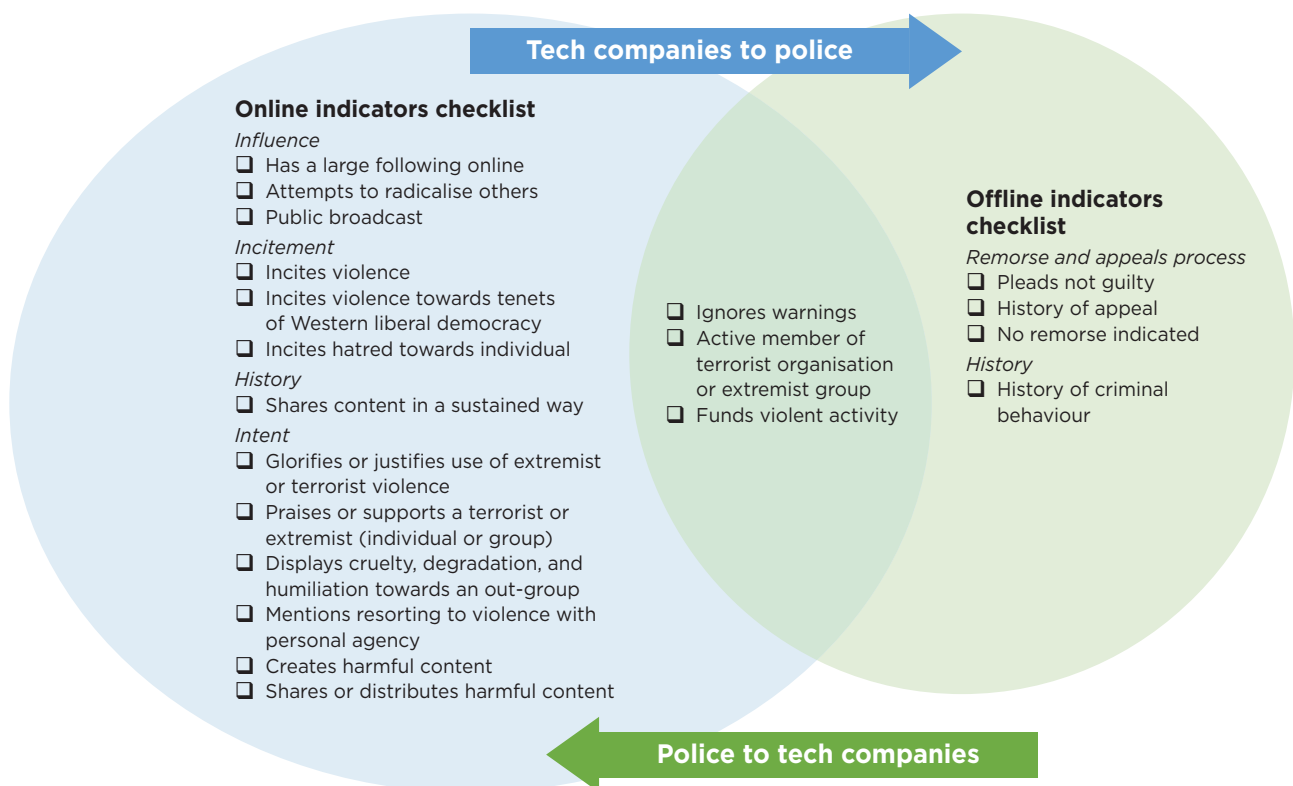


Figure is author's own work

Technology companies should therefore track whether individuals in question meet a threshold of warning signs regarding the 20 proposed indicators, while working closely with the police and with open source court documents to understand patterns of behaviour that may be occurring offline, but which lend themselves to harmful behaviour online (see Figure 3).¹⁶³ It is hoped that these indicators will lead to better consistency across platforms, and better transparency

¹⁶² On social media policies that address extremism, see, for example: 'Hate speech', *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/hate_speech, last visited: 17 December 2019; and 'Violence and graphic content', *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/graphic_violence, last visited: 17 December 2019.

¹⁶³ There are, of course, privacy and scalability issues with such an approach.

between technology companies and offline legislation, in understanding which extremist speakers require close monitoring or removal of privileges following harmful behaviour.

Using a sample of 107 offenders from 2015–2019 who were convicted for various terrorist and/or hate crimes offences with an online component (see Section 5, ‘The Dataset’), offenders were given a ‘1’ if their behaviour matched the extremism indicator below, a ‘0’ if it did not, and an ‘N/A’ if information was not available. The scale was broken down into six ‘harm’ levels, each with a dispersion of approximately three points, in order to best fit profiles which contained grades with decimal places.¹⁶⁴ The results were then mapped on an extremism grading scale from 0–15.2 (with 15.2 representing the highest level of harm online).

It is important to note that although the majority of indicators occur online (13 out of 20 indicators), some behaviour occurs offline (four indicators regarding the appeals process and criminal behaviour) and some activity is a mix of both online and offline spaces (three indicators regarding funding, warnings, and membership). The indicators are outlined below and are highlighted in blue if they are online, in green if they are offline, and in grey if they are a combination of online and offline behaviour.

3.1. Categorisation of 20 indicators

Influence

An important consideration in the five case studies was the influence of a speaker, particularly if that speaker was in a position of power or authority. This was evidenced in the case of FM, who occupied a position as an imam, irrespective of the size of his audience or the ability of this audience to go elsewhere if they did not agree with his views. This has implications for social media companies who may need to moderate individuals on the basis of their influence, as well as, or in addition to, the content of their words. Recent research published by the Global Research Network on Terrorism and Technology has illustrated that bans on individuals and organisations by Facebook have significantly reduced their influence, despite moving to alternative platforms such as Gab – with a reduction in one instance from 1.8 million followers on Facebook to 11,000 followers on Gab.¹⁶⁵ Therefore, social media platforms have a role to play in moderating the influence of speakers and reducing their exposure to users.

Three indicators were created to map the influence of the extremist individual or organisation online. These are:

Has a large following online (content is liked and re-shared; there are many views of, and comments on, the content)

- This indicator was scaled according to number of likes, shares or followers:
 - More than 100 likes, shares or followers was classified as a 1 (see *R. v Gary Staples*)
 - 50 likes, shares or followers was classified as a 0.5.
- The number of likes and shares of the most popular post was taken into account and scaled accordingly, for example, the most popular prayer post shared by an individual, which gained five shares and 14 likes, would be 0.019 (see *R. v Adam Paul Wyatt*).

¹⁶⁴ Harm level categories were of a different size as the inclusion of decimal places in the indicators made it difficult to disperse the categories evenly. In particular, 19 indicators could be graded as a 0, 0.5, or 1, and one indicator ‘has a large following online’ was scaled based on the number of likes, comments, views, and shares of posts.

¹⁶⁵ Lamb, H., ‘Mainstream social media bans “erode influence of extremist groups”’, *Engineering & Technology*, 8 July 2019, available at: <https://eandt.theiet.org/content/articles/2019/07/mainstream-social-media-bans-erodes-influence-of-extremist-groups/>, last visited: 17 September 2019.

- This indicator was expanded to include number of views:
 - 1,000 or more views was classified as a 1
 - 500 views and shares was classified as a 0.5.
- The term ‘following’ included social media linkages such as providing a gateway that enabled fighters to register social media accounts (see *R. v Rabar Mala*).
- Where information was absent about the extent to which the content was liked or re-shared, but the CPS noted that the defendant’s posts regularly received ‘likes’ or ‘shares’, a 0.5 was given (see *R. v Nigel Pelham*).
- Offenders were given a 1 when the extent of the defendant’s online presence was so exceptional that it was very likely that s/he had a large following (see *R. v Alaa Esayed* where her Twitter and Instagram accounts were acknowledged by Al-Qaeda for their outreach ability).
- Cases with an online presence, but where the extent of mobilisation was unknown, were coded 1 under the separate indicators ‘Creates harmful content or ‘Shares or distributes harmful content’ (see *R. v Naseer Taj*).

Attempts to radicalise others (exerting influence)

- Defined as communications involving an online component, conducted between the individual and specific members of the general public, to influence the latter to adopt extreme political, social, or religious ideals or aspirations.
- While thought was given to expanding the indicator to include spreading or re-sharing extremist propaganda or literature (see *R. v Akeem Samuels*, where Samuels shared proclamations of IS leader Abu Bakr al-Baghdadi), it was determined that sharing of violent or non-violent rhetoric or imagery was already encapsulated by the separate indicator ‘Shares or distributes harmful content’. To ensure there was no significant overlap between the two indicators, the indicator ‘Attempts to radicalise others’ captured personalised communication between the defendant and specific members of the general public.

Public broadcast

- Defined as using specific channels on social media platforms that are readily accessible to the general public to upload content with the intent to provide information on, or support for, proscribed organisations (see *R. v a Youth (Operation Groop)*). This definition did not include repeated comments uploaded on YouTube content (See *R. v Abad Ali (Darren Glennon)*), which would fall under the indicator ‘Shares content in a sustained and deliberate way’.
- For the purposes of this study, social media platforms such as YouTube, Twitter, and Facebook were coded as 1 due to the easily accessible nature of their content, and alternative social media platforms such as Telegram, VK and WhatsApp were coded as 0, unless court documents specified that material on the former platforms had been purposefully restricted.
- There were several cases where the private nature of the posts was contested by the accused (see *R. v Nigel Pelham*). In such cases, the Counter-Terrorism Division of the CPS’s post-conviction summary analysis of the posts was used to determine whether the indicator applied to the individual in question.

Remorse and Appeals Process

What was clear from *Naik* was the ability of an offender to refute the claim that they are an extremist, by allowing the individual in question to submit evidence that their views had changed. Such remorse, however, had to be balanced with the risk that the speaker

could marginalise the significance of their statements, or use semantics to understate their significance. Therefore, the following four categories were created to map a display of remorse or a justification refuting claims that an individual is extremist, potentially leading to re-evaluation of an individual's status as an extremist capable of causing harm.

Pleads not guilty

- Defined as the individual pleading not guilty to the offence in question.
- Under this indicator, 0 would indicate pleading guilty, 1 would indicate pleading not guilty, and an N/A tag would mean information was absent about the defendant's plea.

History of appeal

- Defined as the individual appealing against the decision to a judge in a higher court or tribunal.
- Under this indicator, 0 would indicate no history of appeal and 1 would indicate history of appeal.
- An appeal was understood as being absent unless otherwise specified on the CPS database (see *R. v Forhad Rahman*, *R. v Adeel Brekke* and *R. v Kaleem Kristen Ulhaq*, where it was specified that Ulhaq is appealing his sentence).

No remorse indicated

- Defined as the individual showing a lack of remorse, repentance, or an acknowledgment of gravity of harm during their trial or conviction proceedings.
- Under this indicator, 0 would indicate remorse, 1 would indicate no remorse, and an N/A tag would mean information was absent about the individual's remorse or sense of guilt.

Ignores warnings from technology companies, friends, members of the public, or police

- Defined as the individual having previously received a warning from technology companies, friends or police personnel about engaging in the creation or dissemination of online extremist content, but continuing to publish said content (see *R. v Aabid Ali aka Darren Glennon*, *R. v Andrew Littlefair* and *R. v Shane Fletcher*). This indicator was also coded as 1 if the defendant had a prior history of radicalisation, for instance, having been visited or given warnings by officers in the Prevent programme.
- Under this indicator, 0 would imply no history of ignoring prior warnings from technology companies, friends, or police, 1 would imply a history of ignoring prior warnings, and an N/A tag would mean information was absent about the defendant's history of warnings.

Incitement

A key theme in the five qualitative case studies was the ability of the individual to incite violence or risk inciting violence through their speech, as well as their proximity to the centre of potential violence. The balance between security and the right to receive information was especially difficult when it came to issues of extremism and, as such, social media companies would need to work closely with police and courts to understand which extremist speakers are problematic. Nonetheless, three categories were created to examine the trend of incitement amongst extremist speakers online:

Incites violence

- Defined as any rhetoric or depiction that encourages individuals to undertake violent acts in support of a cause (see *R. v Gary Staples*, and Staples' explicit typed video slides

superimposed with the words “Come to Jihad”; *R. v Akeem Samuels* with Samuels’ images with the message, “Shi’ism – cut it out before it spreads”).

- This indicator included distributing recordings of audio or visual images encouraging listeners to undertake violent acts, although the persons involved in making or being part of the recording may or may not be the person distributing the material in question (see *R. v Nourdeen Abdullah Al-Gharib*).
- This indicator would include the delineation of explicit methods of enacting violence, for example, through the dissemination of a manual describing ways to kill members of one racial group (see *R. v Lawrence Burns*), even though there was no specific injunction to kill.
- This indicator would not include explicit references to specific terrorist attacks in the past, which would be categorised under the indicator ‘Glorifies or justifies the use of extremist of terrorist violence, or a past act of extremist of terrorist violence’.¹⁶⁶

Incites violence based on antagonism towards the tenets of Western liberal democracy and the armed forces

- Defined as any rhetoric or depiction that encourages individuals to undertake violent acts against the armed forces, the West or existing political order.
- This would include explicit references to alleged atrocities committed by the West or the promotion of beliefs or narratives that delegitimise existing national or international order, with the specific intention to provoke others to commit violent acts, or to undermine or not cooperate with Western institutions (not voting, for instance, or likening police forces to the devil).
- Instances involving IS propaganda were examined on a case-by-case basis (due to the existence of non-violent propaganda featuring IS branding).

Additional considerations:

- The indicator ‘Incites violence based on antagonism towards the tenets of Western liberal democracy and the armed forces’ was derived from the UK government’s existing definition of extremism, despite the essentially contested nature of what ‘the tenets of Western liberal democracy’ constitutes. Nonetheless, used in this context, the indicator was meant to encapsulate a threefold set of rights:¹⁶⁷ (i) negative freedoms, including the freedom of conscience, a free press and free speech, equality under the law, and the right to hold and exchange property without fear of arbitrary seizure; (ii) positive freedoms, the rights necessary to protect and promote the capacity and opportunity for freedom, and (iii) democratic participation or representation. It is acknowledged that this indicator relies on a largely Ameri-centric and Euro-centric definition of ‘the tenets of Western liberal democracy’.
- Consideration was given to altering the indicator to capture the incitement of hatred, rather than of violence. However, cases featuring hatred were likely to be captured under the indicators ‘Displays cruelty, degradation, and humiliation towards an out-group’, or ‘Incites hatred towards a particular individual or several individuals in the interests of discriminatory harm’. It was important, moreover, to ensure that legitimate criticism of and opinion about political parties and against the West was not captured under this indicator (in order to preserve the ability of people to have free speech and opinions on political issues online, even if they veer towards ‘hatred’).
- Unlike Islamist groups, far-right organisations tended to glorify the West or Britain and use derogatory and racist terminology to predict its demise (including the use of theories of

¹⁶⁶ This indicator did not include indirect and/or unintentional encouragement of violence.

¹⁶⁷ Doyle, M. W. ‘Kant, Liberal Legacies, and Foreign Affairs’, *Philosophy and Public Affairs*, 12. 3 (1983): pp.205-235; pp.206-207.

white replacement and mass immigration).¹⁶⁸ A case in question would be *R. v Jennings*, where Jennings, a self-described “libertarian” and “proud political dissident”¹⁶⁹ had made threats against Muslims, Jews, and members of the Labour Party. The potential that this indicator would downplay far-right groups was mitigated by the presence of alternative indicators, including ‘Incites violence’ and ‘Displays cruelty, degradation, and humiliation towards an out-group’, which would include far-right behaviour.

Incites hatred towards a particular individual or several individuals in the interests of discriminatory harm

- Defined as singling out and using harmful language, slurs or insults towards an individual or several individuals based on their personal characteristics or affiliations.

History

Apparent from both *Salah* and *Naik* was the need for the court to consider weighing isolated extremist statements on a single occasion versus a history of extremist behaviour. In the online space, such an approach would need to be factored into the historical profiles of those with large followings online, who may violate a platform’s three strikes policy. Therefore, three indicators were created to reflect the history of the speaker in question, to grade their level of extremism.

Shares content in a sustained and deliberate way (intent); illustrates multiple instances of sharing harmful content (history)

- Defined as multiple repeated instances of rhetoric, recordings or supportive imagery of proscribed organisations (including the use of symbols, flags, banners, and so on).

History of criminal behaviour

- Defined as a previous history of criminal offences convicted under the law, including, but not limited to, drug-taking, petty theft, burglary, child cruelty, and/or domestic violence.
- This indicator would not include a previous history of sharing harmful content or expressing online support or supportive imagery, which would be captured under the indicator ‘Shares content in a sustained and deliberate way’.

Active member of a proscribed terrorist organisation or actively associated with an extremist individual or group

- Defined as having declared allegiance for, and been recognised by, proscribed terrorist organisations, or as being part of said organisation.
- Instances involving self-declared supporters of organisations, but whose membership was not formally recognised by the group, were coded as 0.5. Such cases included individuals who shared violent or non-violent online content, but whose actions in the online sphere were not mirrored by personal actions demonstrating support in the offline realm (such as travel or attempted travel to join the proscribed organisation, or attendance at offline meetings; see *R. v Mohammed Khan*, where Khan reposted a tweet of an IS call to attack US bases in Bahrain). Instances involving expressed declarations of support mirrored by personal actions demonstrating exceptional support and commitment to the organisation,

¹⁶⁸ Allen, C., ‘National Action: links between the far right, extremism and terrorism’, University of Leicester (2019), available at: <https://www.gov.uk/government/publications/national-action-links-between-the-far-right-extremism-and-terrorism>, last visited: 13 August 2019, p.8.

¹⁶⁹ Evans, J., ‘Jail for man who posted “vile and venomous” social media messages saying Muslims should be gassed’, *Wales Online*, 9 August 2018, available at: <https://www.walesonline.co.uk/news/wales-news/jail-man-who-posted-vile-15008433>, last visited: 7 August 2019.

such as travel or attempted travel to join the organisation (see *R. v Gary Staples*) or the setting up of a self-help online library for terrorists (see *R. v Samata Ullah*) were coded as 1.

- This indicator was primarily created to categorise self-declared supporters of far-right organisations; the individual's presence at the meeting of a far-right organisation would be tagged as indicative of being an active member of an extremist group (see *R. v Alison Chabloz*).

intent

Finally, an overall theme common to all five qualitative case studies of extremism was the need to balance the right of an audience to receive information, versus protecting an audience from extremist views or restricting this information due to interests of national security, public safety, prevention of disorder and crime, and protection of health and morals. There was also a risk that individuals whose views are not extremist, but deemed to be so based on information at the time, may be banned from the platform. Moreover, individuals who criticised religion in an opinionated way could also be regulated. Therefore, six remaining indicators were created to reflect intent, relying largely on the content linked to each individual.

Glorifies or justifies the use of extremist or terrorist violence, or a past act of extremist or terrorist violence

- Defined as glorification of violence against a protected group for extremist aims (such as the Holocaust) or validation or justification of specific terrorist attacks such as the September 11 attacks or the Lee Rigby attack (see *R. v Akeem Samuels*).
- It was acknowledged that the indicator 'Glorifies or justifies a past act of extremist or terrorist violence' may overlap significantly with the indicator 'Praises or supports a proscribed terrorist organisation or individual'. However, both indicators remain, as the celebration of violence is a significant element of extremist propaganda, and should not be subsumed under the more expansive category 'Praises or supports a proscribed terrorist organisation or individual' (see *R. v Lloyd Gunton*).
- This indicator is different from 'Incites violence', which would include more targeted rhetoric or depictions calling for individuals to directly enact acts of violence.

Praises or supports a proscribed terrorist organisation or extremist (individual or group)

- Defined as publicly expressing support for, or publishing violent or non-violent imagery for, a proscribed terrorist organisation and/or extremist individual or group. For example, this may include sharing symbols, flags, and banners of a proscribed terrorist organisation outlawed in the UK.¹⁷⁰ This indicator may also include praise or glorification of extremist organisations, such as the Nazi Party, or Holocaust denial, which is not protected from restrictions against hate speech online that criminalise intentionally 'grossly offensive' material (see *R. v Alison Chabloz*).
- This indicator includes rhetoric or depictions of militants in combat in a supportive manner, such as links to stories in the news of IS poisoning wells in Syria (see *R. v Amir Maqbool*), as opposed to news content or content uploaded for the purpose of exposing and condemning such violence.

Displays cruelty, degradation, and humiliation towards an out-group

- Defined as any rhetoric or depiction involving cruelty, degradation or humiliation enacted specifically against minority groups due to their religion, race, nationality (including citizenship),

¹⁷⁰ 'Proscribed terrorist groups or organisations', HM Government (2013), available at: <https://www.gov.uk/government/publications/proscribed-terror-groups-or-organisations--2>, last visited: 13 August 2019.

sexual orientation or other physical, cultural or behavioural characteristics, including but not limited to communities belonging to the Jewish or Islamic faiths, intra-Muslim groups and members identifying as being part of the LGBTQI+ community, amongst others.

- This does not include incitement of violence against states including the US and Israel (see *R. v. Nourdeen Abdullah Al-Gharib*), which would be captured under the indicator ‘Incites violence based on antagonism towards the tenets of Western liberal democracy and the armed forces’.

Funds violent activity

- Defined as money or property making, or likely to make, a significant contribution to furthering terrorism, or the use or provision of money or property to fund or assist activities endangering life, causing widespread or serious damage to property or economic interests, or substantial impact upon civic infrastructure.
- This indicator may or may not involve self-declared support of a proscribed organisation (see *R. v Salim Wakil*).
- The resulting conviction also contributed to this indicator, where the individual was convicted under terrorism-funding legislation.
- Serious consideration was given to using the phrase ‘funds terrorist activity’ to describe this indicator. This was ultimately replaced by ‘funds violent activity’ due to the latter term’s wider inclusiveness. The tag ‘funds violent activity’ was able to account for the funding of racially or religiously motivated violent crimes unaffiliated with a proscribed organisation (see *R. v Rhodri Philipps*).

Mentions resorting to violence with personal agency (“I want to kill...”)

- Defined as making explicit threats online involving a personal element to engage in a terrorist act (see *R. v Lloyd Gunton*), or expressing an intent to assist in acts of violence for extremist or terrorist purposes (see *R. v Gary Staples*).
- This would include attempts to join a proscribed organisation such as IS, but only if such attempts included an online component. As such, cases such as *R. v a Youth (Operation Groop)* would not qualify under this indicator, because an online component was lacking before or during the act.
- This indicator would include personal rewards offered for violent acts (see *R. v Rhodri Philipps*).
- Due consideration was given to expanding this indicator to include calls to violence without the personal agency element (see *R. v Nigel Pelham* or *R. v Andrew Emery*, where Emery posted a message that “we should burn a mosque”). However, it was concluded that this indicator should focus only on posts directly involving the first-person pronoun to capture a level of personal responsibility, and references to a collective group were encapsulated by the indicator ‘Incites violence’.

Creates harmful content

- Defined as the creation of original extremist violent or non-violent content within the online sphere, specifically rhetoric, video or audio footage of violence, armed conflict or similar, in a manner that could be deemed to be supportive.
- Much consideration was given to this indicator to distinguish between alleged ‘keyboard warriors’ (individuals who out of carelessness or inadequate forethought may have posted harmful content online; see *R. v Andrew Emery*) and individuals who exhibited demonstrably higher levels of effort and commitment towards content creation (see *R. v*

Gary Staples, who created eight videos combining still photographs, nasheed songs, and typed message slides). Intended harm in the latter case was assessed to be demonstrably greater, due to the higher level of forethought and commitment invested in the cause (see *R. v. Saer Hussain*, where the CPS noted that an existing mitigating factor for his plea was that Hussain “did not create material but [only] relayed material created by others”¹⁷¹).

- Converted to a numerical coding, this indicator was scored on a three-point system: 1 (present), 0 (absent) and 0.5 if the individual did create some original content, but this was largely in the shape of personal opinion (see *R. v Andrew Littlefair*) or the content comprised a small part of the individual’s modus operandi (see *R. v Lloyd Gunton*). Cases which involved rhetorical content and the posting of images (see *R. v John Hanson* where Hanson uploaded a photo of himself with a t-shirt and the words “destroying Islam”) were coded as 1.
- This indicator would include participation in the creation of harmful content, for example, the defendant giving a speech with the knowledge that it would be filmed (see *R. v Lawrence Burns*).

Shares or distributes harmful content

- Defined as the republication of harmful rhetoric or depictions created by others, but shared in a manner that could be deemed to be supportive.

¹⁷¹ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016’, *Crown Prosecution Service Counter-Terrorism Division*, undated, available at: <https://www.cps.gov.uk/counter-terrorism-division-crown-prosecution-service-cps-successful-prosecutions-2016>, last visited: 13 August 2019.

4. The Dataset

Between 6 June and 10 July 2019, six FOI requests were made to the CPS, asking for information on how many individuals were prosecuted for ‘violent extremism’, offline and online, between January 2009 and January 2019. The requests also asked how many individuals had been prosecuted for hate crimes with a social media element in a shorter time frame between January 2015 and January 2019.

All of the requests were refused except for one, which did not receive a reply. A response from the CPS dated 20 June 2019, however, revealed that ‘violent extremism’ was not considered a stand-alone offence that could be convicted, and is instead a behaviour that could be attributed to several offences. The CPS indicated that determining the number of cases in which acts of violent extremism occurred in the last five years would involve examining all offences with a ‘hate crime’ component, moreover, counter terrorism statistics would also be used. Extremism was prosecuted on the grounds of both terrorism and hate crime.¹⁷²

The UK government’s counter-extremism strategy identifies terrorism and hate crime as the results of extremism.¹⁷³ As such, a hybrid approach examining publicly available hate crime offences and terrorist offences between 2015 and 2019 was employed to put together a database of 260 offenders. Individuals were identified as being extremist if:

- They were convicted for terrorism offences in a British court, and the case was publicly available under the counter-terrorism division of the CPS website.
- They were prosecuted under non-terrorism legislation but the offence could be matched to the UK government’s definition of extremism.¹⁷⁴ Typically, this included offences contrary to common law, such as murder, or public order offences such as incitement to hatred based on race, religion, or sexual orientation. These cases were found primarily through the Community Security Trust (CST) open source log of cases and were collectively referred to as “extremism-related”. Additional cases were found by reading the annual reports released by Tell MAMA from 2015–2018. Three cases overlapped between CST and TellMAMA during this period.

Due to the focus of this paper on the online space, the cases were subsequently examined for an online element, meaning use of popular social media platforms and/or encrypted apps to spread extremist views. Of the 260 original cases, 107 cases were deemed to contain an online element. Of these:

- 79 cases were found solely on the CPS database
- 10 cases were found only on the CST case log.
- 11 cases were both CPS and CST cases, publicly available on both the CPS website and the CST website
- 2 cases were both CPS and TellMAMA cases, publicly available on both the CPS website and in the Tell MAMA annual reports
- 5 cases were from the HJS report *Islamist Terrorism, 1998–2015*

¹⁷² Information Management Unit, Crown Prosecution Service, ‘Re: Number of offences regarding violent extremism’, received by Nikita Malik, 20 June 2019.

¹⁷³ ‘Counter-Extremism Strategy’, HM Government (2015), p.5-6.

¹⁷⁴ The UK government defined extremism in its 2011 revised ‘Prevent’ strategy as “the vocal or active opposition to fundamental British values, including democracy, the rule of law, individual liberty and mutual respect and tolerance of different faiths and beliefs. We also include in our definition of extremism calls for the death of members of our armed forces, whether in this country or overseas.” ‘Prevent Strategy’, HM Government (2011), available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/97976/prevent-strategy-review.pdf, last visited: 13 August 2019, p. 109.

Where there was confusion as to whether a case should be included, policies published by social media platforms on ‘harmful or dangerous content’, ‘hateful content’, and ‘violent and graphic content’ were also used.¹⁷⁵

The 107 publicly available cases involving hate crime and terrorism online from 2015–2019 were then coded using the 20 extremism indicators outlined above. Several rounds of coding were done by two researchers to ensure that indicators reflected common patterns amongst those at different grades of extremism. Although an inter-coder reliability score was not used, differences where apparent were discussed and resolved.

4.1. Limitations

This study has made use of the CPS database which, while extensive, is not exhaustive. Data on the CPS database relating to cases involving high-profile suspects is often limited, redacted, or not reported due to security concerns. To mitigate this limitation, the study has made substantial use of additional media reports and local news sources to corroborate and substantiate information available on the CPS database. When adequate information was not available on the individual, the profile was not included. As a result, the cases used in the extremism grading framework below are illustrative, rather than comprehensive.

It is important to note that the resulting extremism grading framework primarily measures harm as a result of extremist activity on social media platforms, rather than harm from disrupted terrorist activity. Therefore, individuals may have high sentencing under existing counterterrorism legislature, but possess a low level of extremism on the indicator index. An example is *R. v Salim Wakil*, where Wakil received a sentence of 30 months’ imprisonment, but his low extremism indicator score of 2 (one of the lowest in the database) reflects his low social media presence and absence of public broadcasts or sharing of extremist beliefs online. Rather, the framework indicates an assessment of online harm, as perpetuated by individuals such as Husnain Rashid, who had an extensive social media presence and engaged in prolific sharing of dangerous, hateful, and violent content, resulting in a high grade of 11.5 for this case.

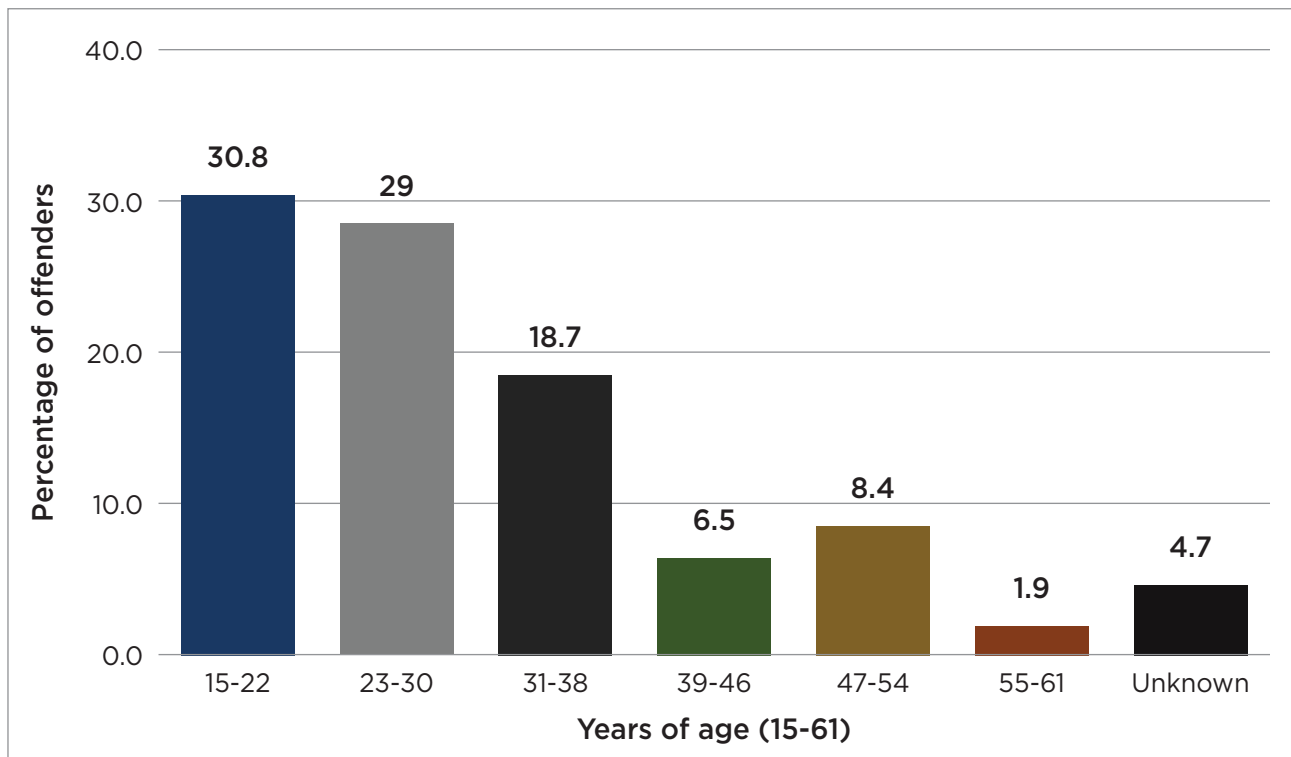
¹⁷⁵ YouTube, in particular, has enacted hate speech policies to encompass “content promoting violence or hatred against individuals or groups based on any of the following attributes: age, caste, disability, ethnicity, gender identity, nationality, race, immigration status, religion, sex/ gender, sexual orientation, victims of a major violent event and their kin [and] veteran status.” ‘Hate speech policy’, *YouTube*, 5 June 2019, available at: <https://support.google.com/youtube/answer/2801939?hl=en>, last visited: 19 June 2019. Facebook has defined hate speech as “a direct attack on people based on what we call protected characteristics – race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity and serious disease or disability.” ‘Hate speech’, *Facebook*, available at: https://www.facebook.com/communitystandards/objectionable_content, last visited: 19 June 2019.

5. Extremism Harm Framework: Scaling Harm

Between 2015 and 2019, 107 cases within the UK where the judge used various offences (terrorism acts, religious and racial hatred, incitement, malicious communication) to police extremism online were examined for effectiveness, using 20 indicators to grade levels of extremism. Different recommendations are suggested based on common patterns within each level of harm (from a 0-15.2 scale, where 15.2 represents the greatest level of extremist harm online).

5.1. Background of Offenders

Figure 4: Age at Conviction



As illustrated in Figure 4, offenders using online platforms for extremist purposes tended to be young, with the most common age at conviction being 15-22 (30.8%) followed by 23-30 (29%). Of the 107 cases examined, 87% were male and 13% were female. 21% of offenders were employed, 7% were unemployed, and data was unavailable for the majority (65%).

In terms of location, Figure 5 shows that the majority of offenders were based in London, at 34 offenders (31.7%), followed by North West England (13%), and West Midlands (12.1%). It was not possible to ascertain the background of the offenders in terms of nationality and ancestry, as not enough information was available for each of the cases.

The majority of hate material shared was Islamist (shared by 71.9% of the offenders in the database), followed by material of a far-right nature (shared by 27.1% of offenders on the database). One instance was categorised as ‘other’, that is, material promoting the Partiya Karkeren Kurdistanî (PKK).

As shown in Figure 6, the platform used most by extremist offenders in the database was Facebook, which was used by 39 out of the 107 offenders in the dataset. The next most commonly used platform was Twitter, used in 22 cases, followed by WhatsApp, used in 19 cases. It is important to

Figure 5: Geographical Breakdown of Offenders

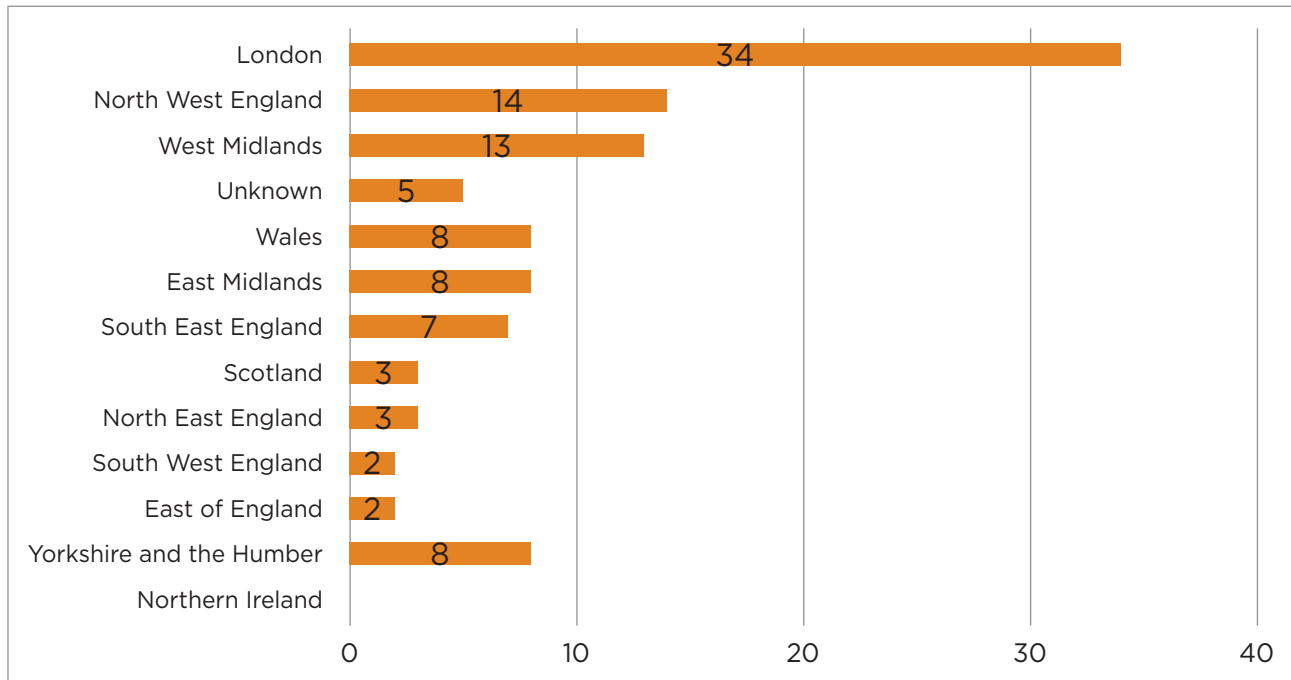
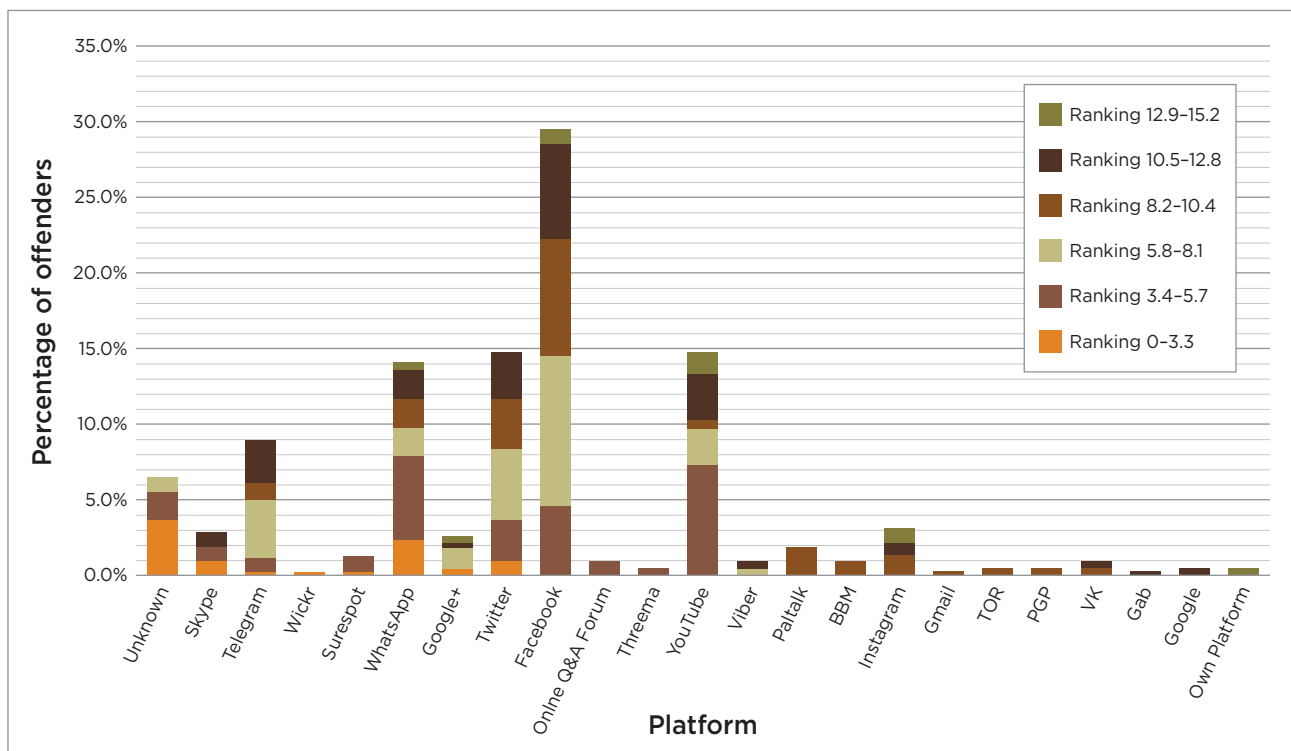


Figure 6: Use of Social Media Platforms by Category Rankings



note that offenders tended to use multiple platforms to communicate, and migrated to platforms such as VKontakte (VK) when they were banned from more popular platforms.

The greatest most common charge faced by individuals in the dataset was terror-related, faced by 75.5% of cases profiled. This was followed by public order-related offences, which were faced by 22% of cases in the dataset. Figure 7 shows how these two charges compare with others. It is important to note that individuals tended to be charged with multiple offences during trial, and care has been taken to update offences with any corresponding appeals where information on the processes was publicly available.

Figure 7: Primary Charge against Offenders by Category Rankings

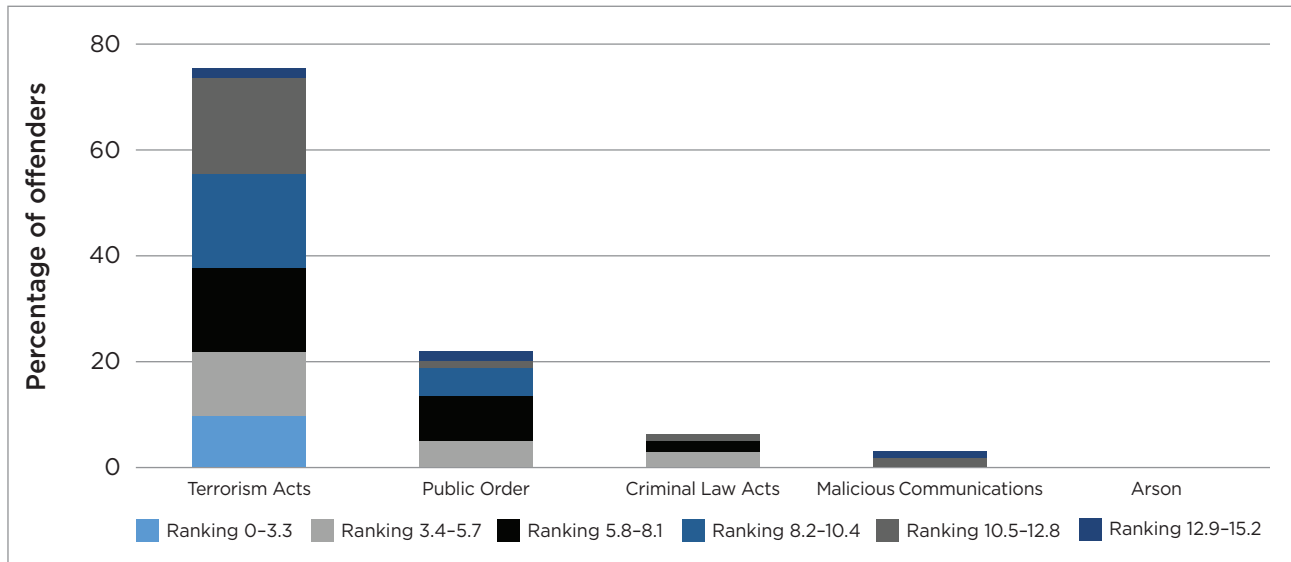


Figure 8: Organisational Affiliation by Category Rankings

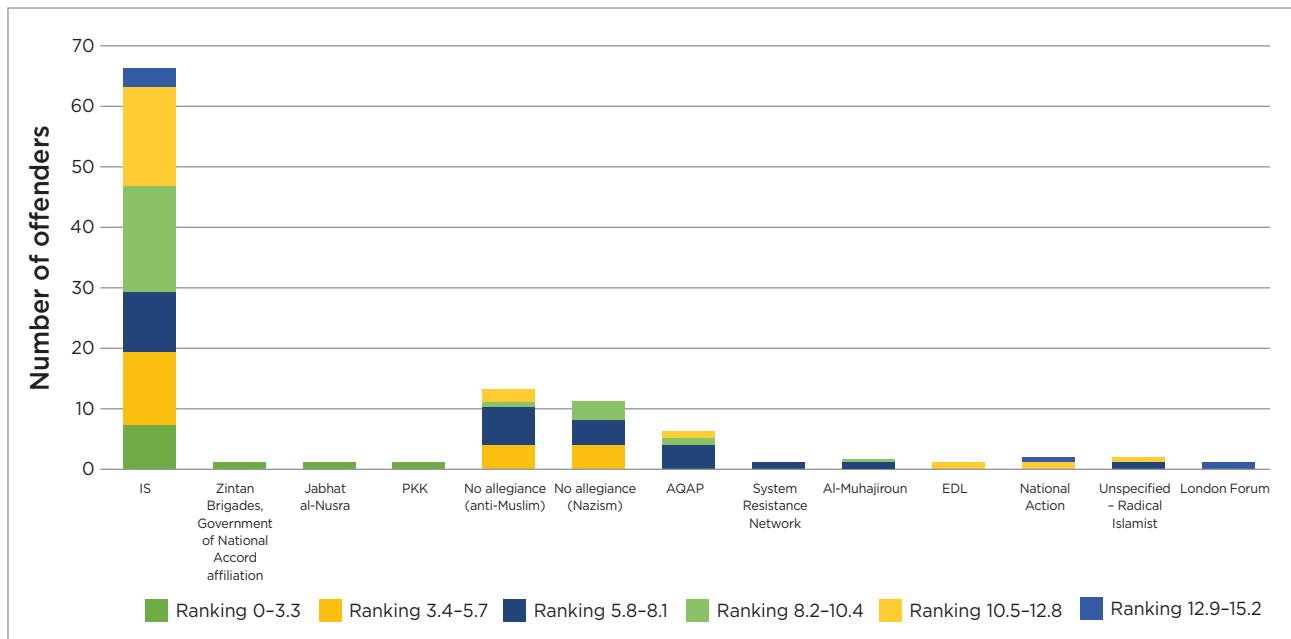
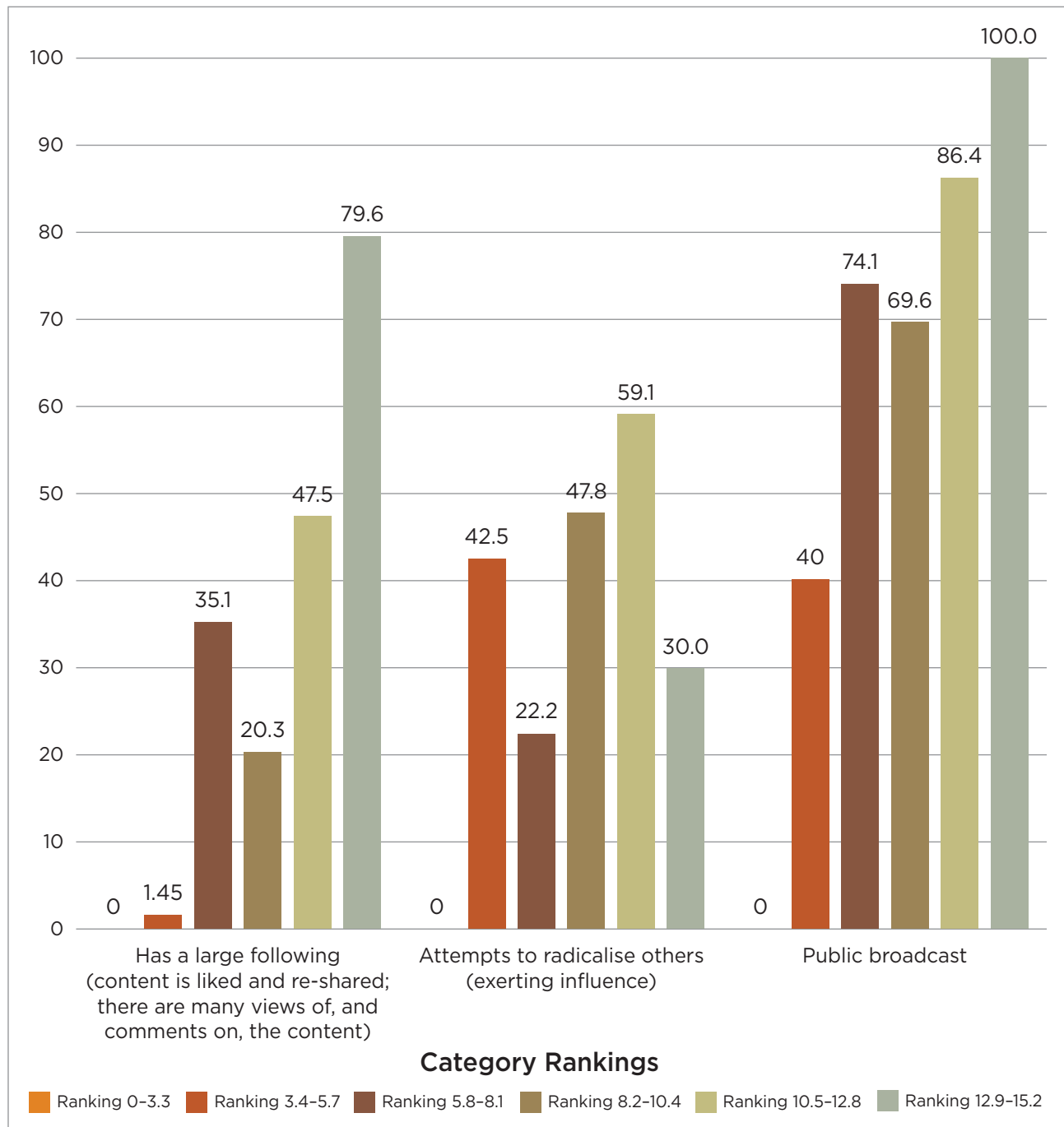


Figure 8 illustrates that the largest organisational affiliation of offenders in the dataset was with IS, affiliated in 61.2% of cases, followed by no affiliation (anti-Muslim) at 12.1%, and then no affiliation (anti-Semitic or Nazi-related content) at 10.2%. 6 of 107 cases in the dataset (5.6%) were affiliated with AQAP. ‘Affiliation’ was qualified as an expression of support, history of involvement with the organisation through previous court records, and any travel plans to join the organisation in question. It is important to note that some offenders were affiliated with multiple organisations.

5.2. Category Ranking: 0-3.3

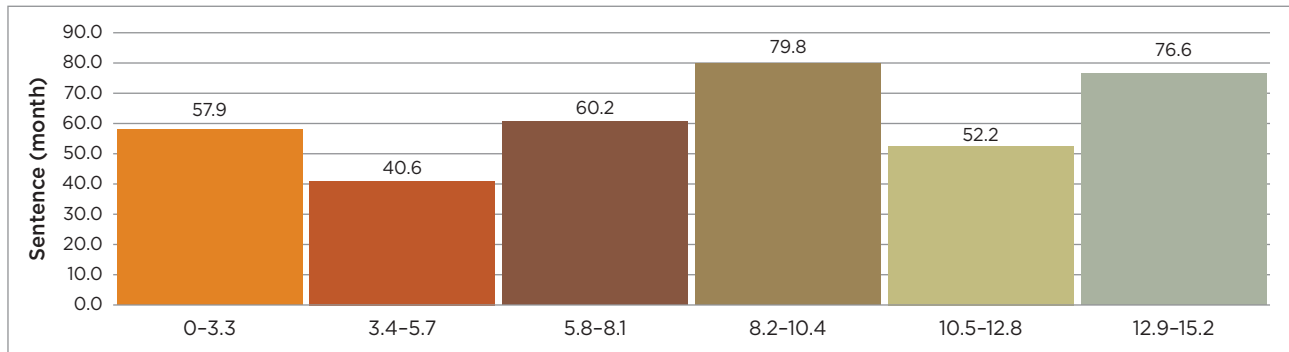
5.2.1. Analysis

The category 0-3.3 is indicative of cases with the lowest level of extremist harm online. As Figure 9 indicates, none of the individuals in this category had high measurements on indicators associated with a prolific online social media presence: they did not have a large online following, there was no content that was liked or reshared, there was no element

Figure 9: Influence Indicators by Category Rankings

of public broadcast, none of the offenders shared or distributed harmful content, and no offenders shared content in a sustained and deliberate way.

However, given extremism is often correlated with, but not subsumed by, terrorism, individuals in this category tended to display high levels of terrorism but low levels of extremist harm online. All offenders in this category were convicted under the Terrorism Act 2000 (TA 2000) or the Terrorism Act 2006 (TA 2006), with 60% under the TA 2000 and 40% under the TA 2006. One explanatory factor for this is that individuals in this category tended to engage in actions in the offline rather than the online sphere. As a result, the case studies profiled in this category would be of particular interest to the police, and data could be actively shared with social media platforms to ensure that profiles are not created online to facilitate terrorist activities. For example, 50% of the individuals in this category funded violent activity.

Figure 10: Average Sentence in Months by Category Rankings

The terrorist-related activity in this category ranking tended to carry higher sentences under the Terrorism Acts 2000 and 2006, with an average imprisonment of 57.9 months, as per Figure 10.

In terms of organisational affiliation, 70% of offenders in this category were associated with IS, 10% were associated with the PKK, 10% with Jabhat al-Nusra, and 10% with Zintan Brigades, a Government of National Accord-funded insurgent group in Libya.¹⁷⁶ This information is shown in Figure 8. However, being an active member or part of a network did not always translate into praising and supporting the proscribed terrorist organisation or an individual extremist online. Shivan Zangana, for example, praised and supported IS online, but Sana Ahmed Khan (associated with Mohamed Rehman)¹⁷⁷ did not praise or support a group or cause online. As a result, technology companies would benefit from surveying court records on individuals who cross a certain threshold of online indicators, and cross-referencing offline behaviour and offences with activity online.

Some individuals were given a 0.5 for the indicator 'Active membership of a proscribed terrorist organisation or actively associated with an extremist individual or group'. Ahmed Ismail, for example, was given a 0.5 as he had changed his mind after making plans to leave the UK to join IS with his friends. Abdulraouf Eshati was also given a 0.5, as he was found to have assisted in an illegal arms delivery to the Zintan Brigades.¹⁷⁸

Offenders in this category were coded highly as being active members of proscribed terrorist organisations, or actively associated with an extremist individual or group. It was interesting to note that cases in this category were drawn into committing terrorist-related offences through mutual encouragement and influence from other extremists. These associations were based on personal networks, such as family members or friends. Salim Wakil, for instance, was motivated by his sister, a member of IS, to transfer money to fund her safe return to the UK,¹⁷⁹ while Majdi Shajira was convicted for conspiring to send a pair of steel-toecapped shoes to his brother, a member of IS.¹⁸⁰

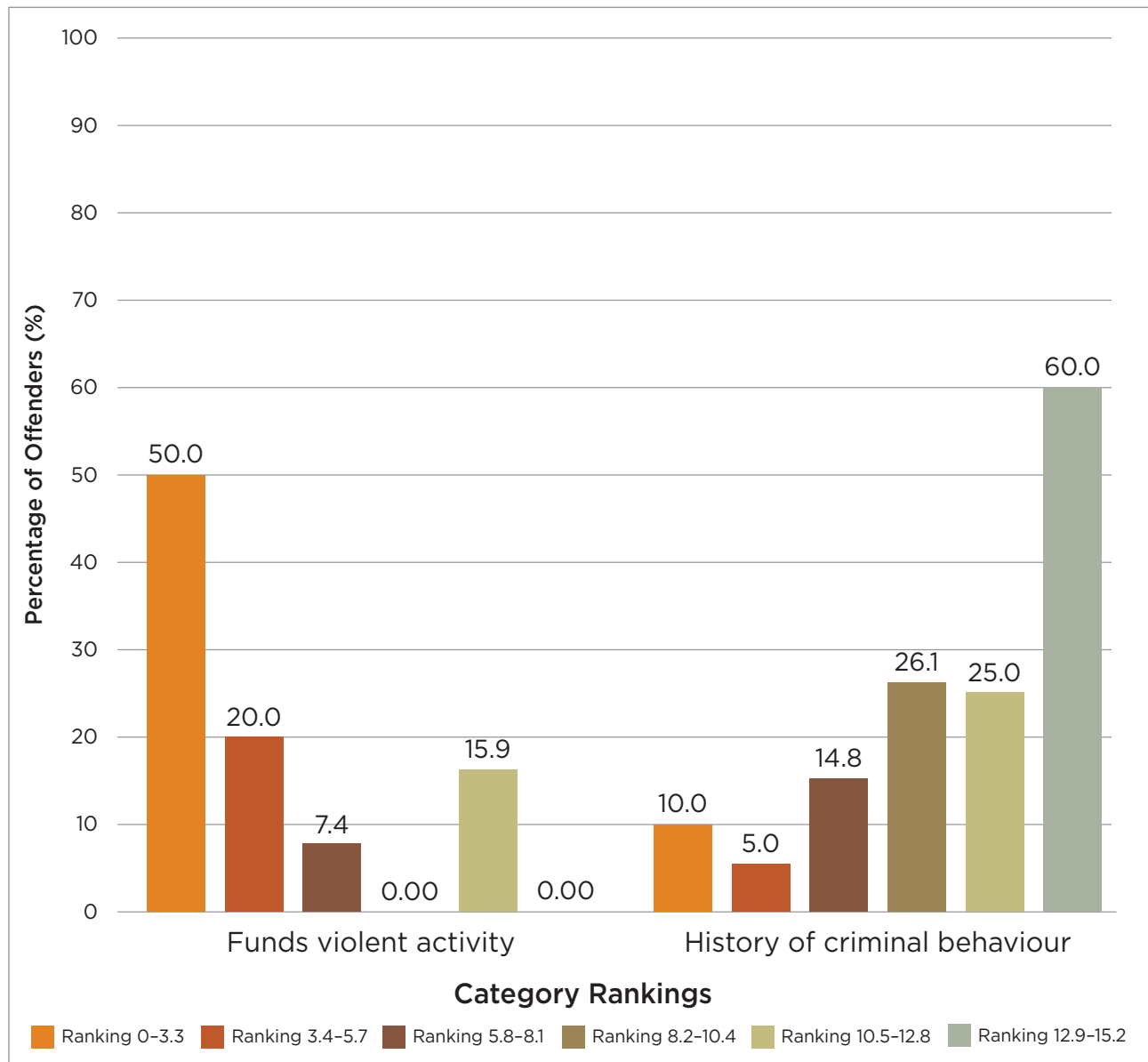
¹⁷⁶ The offender in question was also found in possession of an image of "what appeared to be" a flag of Ansar Al-Sharia, also known as Ansar al-Sharia-Benghazi (AAS-B) which was proscribed in November 2014. However, as affiliation was not confirmed this was not included.

¹⁷⁷ Together with Rehman, known as the "silent" bomber, Khan and Rehman were convicted of plotting a large-scale explosive attack on Westfield shopping centre in London on the tenth anniversary of the 7 July 2005 bombings. Rehman and Khan were married.

¹⁷⁸ 'Wrexham imam Abdulraouf Eshati jailed over Libyan arms plot', *BBC News*, 27 October 2015, available at: <https://www.bbc.co.uk/news/uk-wales-north-east-wales-34645263>, last visited: 15 November 2019.

¹⁷⁹ 'Brother jailed for funding sister who joined Islamic State group', *BBC News*, 8 February 2019, available at: <https://www.bbc.co.uk/news/uk-england-hampshire-47170150>, last visited: 10 August 2019.

¹⁸⁰ Osuh, C., 'Mersey man prosecuted for sending trainers to his terrorist brother', *Liverpool Echo*, 8 May 2015, available at: <https://www.liverpoolecho.co.uk/news/liverpool-news/mersey-man-prosecuted-sending-trainers-9220647>, last visited: 10 August 2019.

Figure 11: History of Criminality and Funding by Category Rankings

30% of this group glorified and justified the use of extremist or terrorist violence online, as well as a past act of extremist or terrorist violence. This indicator was not directly correlated with praising and supporting terrorist or extremist groups. Mashoud Miah, for example, justified the use of violence, telling the jurors that the Syrian people needed help and 'had every right to be defended' even if violence was used.¹⁸¹ However, he did not publicly express support, or publish violent or non-violent imagery for a proscribed terrorist organisation on a social media platform, as encapsulated by the indicator 'Praises of supports a proscribed terrorist organisation or extremist (individual or group)'.

As illustrated in Figure 11, only 10% of offenders in this category had a history of criminal behaviour. Jade Jasmin Campbell had previously been arrested by the police for falsifying information, having applied for a replacement passport when she already possessed one.¹⁸² Only 10% of

¹⁸¹ 'Two men convicted of involvement in funding Syrian extremists', *The Guardian*, 23 December 2016, available at: <https://www.theguardian.com/uk-news/2016/dec/23/syed-hoque-mashoud-miah-convicted-of-involvement-in-funding-syrian-extremists>, last visited: 10 August 2019.

¹⁸² 'Woman jailed for terrorism offence', *Asian Express*, 28 April 2017, available at: <https://www.asianexpress.co.uk/2017/04/woman-jailed-for-terrorism-offence/>, last visited: 9 August 2019.

offenders created harmful content; Abdurraouf Eshati was charged with possession of articles in connection to the commission, preparation, and instigation of acts of terrorism,¹⁸³ having attempted to translate documents relating to an illegal Libyan arms deal.¹⁸⁴ 20% of offenders had a history of appeal, and 20% ignored previous warnings. Salim Wakil sent money to his sister who joined IS despite police warnings that it was against the law¹⁸⁵ and Sana Ahmed Khan ignored warnings from her parents to disassociate herself from Mohamed Rehman.¹⁸⁶

5.2.2. Trends

- Individuals in this category do not use social media for public presence, influence, or incitement. Their online following is low or non-existent, and they do not display cruelty or communicate their intentions to commit violence online. There is largely no public broadcast element, as offenders tend to use encrypted applications: 30% of the individuals used WhatsApp, 10% used Skype, 10% used Telegram, 10% used Twitter, and 40% used unknown forms of online communication.
- Individuals tended to be convicted under the Terrorism Acts 2000 and 2006 for financing. 60% were convicted under Section 17 of the TA 2000, while the remaining 40% were convicted under Section 5 of the TA 2006.
- 30% praised or supported a proscribed group, namely the PKK and IS. 30% glorified or justified use of violence: for instance, Ahmed Ismail and Shivan Zangana both expressed a desire to engage in violent *jihad* in Syria.¹⁸⁷ These two indicators did not overlap.
- 10% created harmful content and 10% had a history of criminal behaviour.
- 20% ignored warnings, 20% appealed.

5.2.3. Recommendations

- The court cases of some, but not all, individuals in this category will be publicly available. Some individuals in this category have prior criminal history and warnings which technology companies can make note of. As many individuals in this category were convicted under terrorism laws, and are active members of proscribed groups or well-known extremist groups, they can easily be banned online, and this ban can be justified. However, an online ban is unlikely to be required, given that individuals in this category do not have a presence on social media or a large following. Nonetheless, social media companies would benefit from coupling research on offline indicators of history and patterns of behaviour with activity online, particularly funding for terrorist purposes.¹⁸⁸
- Based on the case studies in this category, offenders are vulnerable to incitement based on networks of extremist family members or friends. They use encrypted applications such as WhatsApp, rather than public platforms such as Facebook or YouTube, and it is possible that they are doing so because they are consciously monitoring their behaviour. As such,

¹⁸³ 'Five men in court on Syria-related terror and fraud charges', *BBC News*, 14 December 2014, available at: <https://www.bbc.co.uk/news/uk-30477881>, last visited: 9 August 2019.

¹⁸⁴ Waddington-NW, M., 'Wrexham cleric gets six years for £18m illegal arms to Libya deal', *Daily Post*, 27 October 2015, available at: <https://www.dailypost.co.uk/news/north-wales-news/wrexham-cleric-gets-six-years-10341307>, last visited: 9 August 2019.

¹⁸⁵ 'Brother guilty of funding sister who joined Islamic State group', *BBC News*, 15 January 2019, available at: <https://www.bbc.co.uk/news/uk-england-hampshire-46874831>, last visited: 10 August 2019.

¹⁸⁶ Whitehead, T. and Barrett, D., 'Middle class daughter of magistrate who turned to suicide bomb plotter', *The Telegraph*, 30 December 2015, available at: <https://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/12073343/Islamist-fanatic-dubbed-Silent-Bomber-and-wife-guilty-of-77-anniversary-terror-plot.html>, last visited: 10 August 2019.

¹⁸⁷ 'The Counter-Terrorism Division of the Crown Prosecution Service (CPS) - Successful prosecutions since 2016', *CPS*, undated. Separately, praise and support in this way would violate the terms and conditions of many platforms.

¹⁸⁸ See, for example, Facebook's definition of terrorism which looks at behavioural indicators that are primarily based on offline evidence: 'Dangerous individuals and organisations', *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/dangerous_individuals_organizations, last visited: 17 December 2019.

social media companies would benefit from working closely with the police for cases in this harm category.¹⁸⁹

5.3. Category Ranking: 3.4–5.7

5.3.1. Analysis

Offenders tended to display low to medium levels of terrorism by aiding or abetting another individual (usually a close friend or relative) in terrorist-related activities. Individuals displayed low to medium levels of extremism online, but did not have a prolific social media presence. Offenders had the lowest average imprisonment duration (40.6 months) across all categories (see Figure 10).

As per Figure 7, the majority of individuals in this category were charged under the Terrorism Acts 2000 and 2006, with 35% charged under the TA 2000 and 25% charged under the TA 2006. 35% of offenders were charged under the Public Order Act 1986, and 5% of individuals were charged under the Malicious Communications Act. Offenders in this category were largely secondary terrorist offenders, in that they were related to main terrorist offenders, rather than primary terrorist offenders themselves. This was illustrated by high levels of cases (42.5%) of offenders attempting to radicalise others and exerting influence. Offender Y, included in this category, was in a romantic relationship with Offender X (another offender included in the database and graded in the 10.5–12.8 indicator category), and aided and abetted his extremist mind-set and activities via the exchange of more than 2,000 WhatsApp messages a day.¹⁹⁰

Individuals in this category tended to use private social media platforms to communicate, although this was often in conjunction with the use of public social media platforms. The primary platform used was WhatsApp, followed by Facebook and Twitter. Individuals' activities were largely concentrated in the offline sphere, rather than in the online public domain. For example, Mr Gollamaully and his wife, Mrs Gollamaully, sent private messages to their nephew and wired £219 to a Turkey-based IS courier.¹⁹¹ Similarly, Forhad Rahman used Skype to communicate privately with, and assist, another individual, Aseel Muthana, to join IS,¹⁹² while Sandeep Samra used WhatsApp and Telegram to communicate her plans to travel to Syria.¹⁹³

As per Figure 9, individuals in this category did not possess prolific public social media presences. Very few offenders (1.45%) had a large following and published content that was liked and re-shared. Examples of this include Lee Munns, whose anti-Semitic tweet gained four retweets and 23 likes at the time of the capture.¹⁹⁴ In addition, 10% of individuals shared or

¹⁸⁹ Facebook, for example, has law enforcement teams globally. For more, see: 'Law enforcement online request', *Facebook*, available at: <https://www.facebook.com/records/login/>, last visited: 17 December 2019; 'Information for law enforcement', *Instagram*, available at: <https://help.instagram.com/494561080557017>, last visited: 17 December 2019; and 'Information for law enforcement authorities', *WhatsApp*, available at: <https://faq.whatsapp.com/en/general/26000050>, last visited: 17 December 2019.

¹⁹⁰ 'Anzac Day plot: Manchester schoolgirl admits terror offences', *The Guardian*, 26 August 2015, available at: <https://www.theguardian.com/uk-news/2015/aug/26/manchester-schoolgirl-admits-terror-offences-connected-to-anzac-day-plot>, last visited: 10 August 2019.

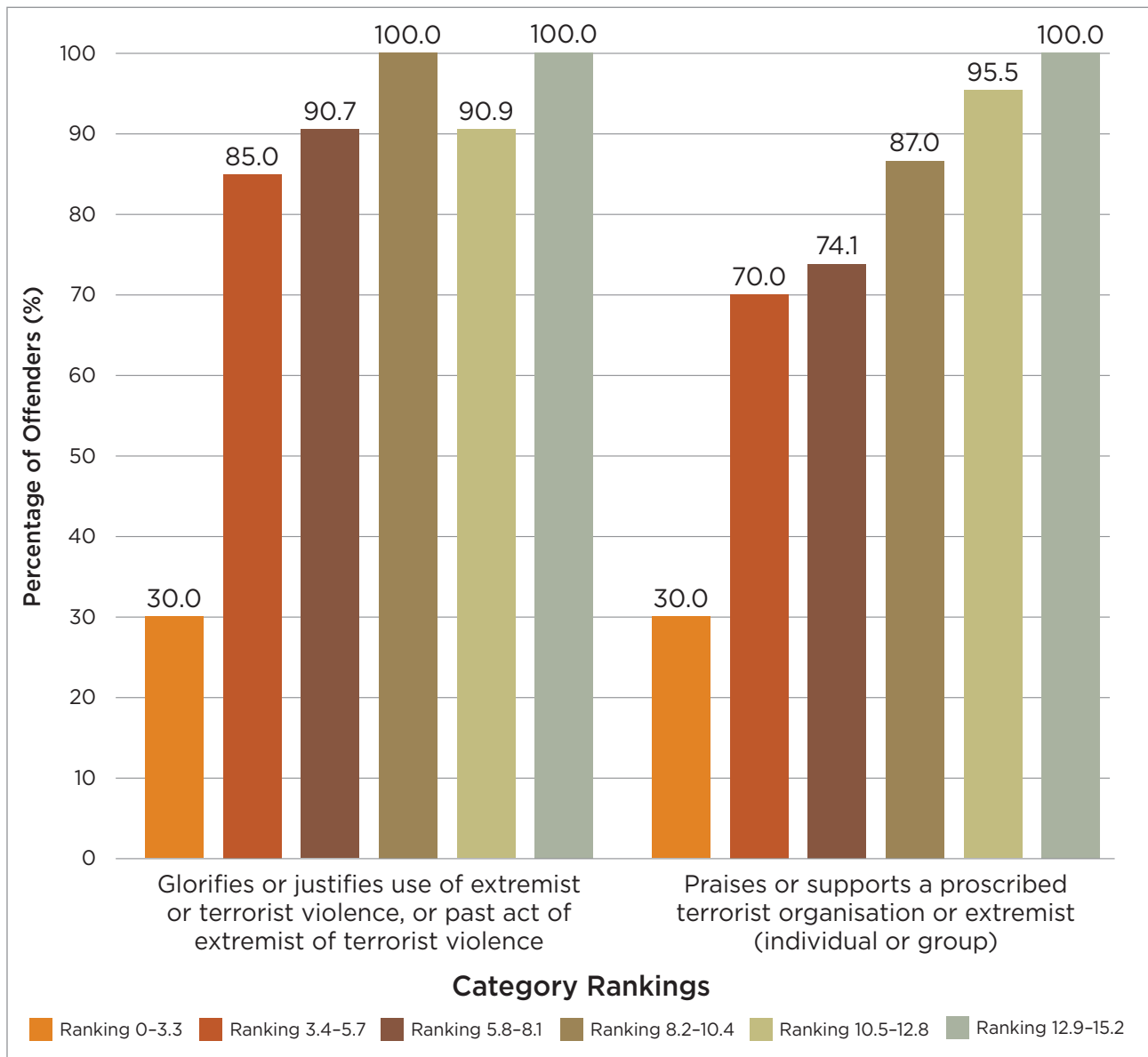
¹⁹¹ 'Muslim extremists who sent money to Isis fighter are jailed for four years', *The Evening Standard*, 22 November 2016, available at: <https://www.standard.co.uk/news/crime/muslim-extremists-who-sent-money-to-isis-fighter-are-jailed-a3401646.html>, last visited: 10 August 2019.

¹⁹² Morris, S., 'Three men found guilty of helping teenage jihadi travel from UK to Syria', *The Guardian*, 10 February 2016, available at: <https://www.theguardian.com/uk-news/2016/feb/10/kristen-brekke-adeel-ulhaq-forhad-rahman-guilty-aseel-muthana-syria-isis>, last visited: 10 August 2019.

¹⁹³ 'Student radicalised online jailed for planning to commit terror attacks in Syria', *ITV News*, 26 January 2018, available at: <https://www.itv.com/news/central/2018-01-26/student-radicalised-online-jailed-for-planning-to-commit-terror-attacks-in-syria/>, last visited: 10 September 2019.

¹⁹⁴ 'Football supporter found guilty of antisemitic tweet', *Community Security Trust*, 4 April 2018, available at: <https://cst.org.uk/news/latest-news/2018/04/04/football-supporter-found-guilty-of-antisemitic-tweet>, last visited: 11 September 2019.

Figure 12: Glorification and Support by Category Rankings



distributed harmful content, for instance, Noamann Ejaz shared nine IS videos with his uncle and one with his friend.¹⁹⁵

As illustrated by Figure 12, the highest indicator in this category was ‘Glorifies or justifies the use of extremist or terrorist violence, or a past act of extremist of terrorist violence’ (85%). However, glorification and/or justification was largely conducted on private, rather than public, channels, feeding into a cycle of online radicalisation. Rebecca Poole, for example, was convicted of possessing a terrorist handbook, and had engaged in an online question and answer forum for advice on joining IS.¹⁹⁶ Shamim Ahmed, another offender included in this category, had viewed pro-IS material on YouTube, but was communicating with another individual on the best place to travel to, and fight for IS, privately.¹⁹⁷

¹⁹⁵ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016’, CPS, undated.

¹⁹⁶ Thomas, J., ‘Terrorist sympathiser told of dreams of joining ISIS and becoming a suicide bomber’, *Liverpool Echo*, 11 October 2016, available at: <https://www.liverpoolecho.co.uk/news/liverpool-news/terrorist-sympathiser-told-dreams-joining-12008387>, last visited: 10 October 2019.

¹⁹⁷ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016’, CPS, undated, available at: <https://www.cps.gov.uk/counter-terrorism-division-crown-prosecution-service-cps-successful-prosecutions-2016>, last visited: 10 August 2019.

The second highest indicator in this category was 'Praises or supports a proscribed terrorist organisation or extremist (individual or group)' (70%). Within this group of offenders, 60% demonstrated praise or support for IS, and 15% demonstrated support for Hitler and/or a Nazi ideology. The remaining 25% demonstrated a more generalised anti-Muslim and/or anti-Semitic sentiment.

11 out of the 20 cases (55%) had an overlap between both indicators: glorifying or justifying the use of extremist or terrorist violence and praising or supporting a proscribed terrorist organisation or extremist individual or group. Six cases (30%) glorified or justified the use of extremist or terrorist violence, without the individual praising or supporting a proscribed terrorist organisation. Offenders in these six cases include Harjinder Singh Athwal, Damanpreet Singh, Parwinder Banning, and Mehul Lodie, who published anti-Muslim posts deemed to be inciting religious hatred and which were threatening in content.¹⁹⁸ However, they did not publicly express support or publish violent or non-violent imagery for a proscribed terrorist organisation and/or an extremist organisation.

32.5% of the individuals in this category were active members of terrorist and extremist organisations. All of this 32.5% had attempted to declare allegiance to, or enacted actions in support of, IS. Cases include that of Aweys Shikhey, who had planned to travel to Syria to fight for IS, and had also discussed launching attacks targeting Jewish people and key public figures in the UK.¹⁹⁹ Mohammed Uddin also engaged in online conversations within social media applications on his media devices to declare his intention to enter Syria for the purpose of joining IS and engaging in combat.²⁰⁰

5.3.2. Trends

- Individuals in this category had a very small online following; as such, while content was posted that glorified violence (85%) and praised extremist or terrorist groups (70%), there was less risk of consumption by other users of the platforms in question. There continued to be a significant use of private messaging platforms such as WhatsApp.
- Only 5% of individuals in this category had a history of criminal behaviour. Shamim Ahmed had been arrested for making online and telephone threats against a London bookshop that stocked Charlie Hebdo magazines.²⁰¹
- 40% of individuals in this category relied on public broadcast. Of these individuals, 62.5% used Facebook, while the remaining 37.5% relied on Twitter.
- 42.5% attempted to radicalise others, such as Mr Gollamaully and his wife, Mrs Gollamaully, who sent encouraging WhatsApp messages to their nephew.
- 40% of offenders incited violence. 45% of offenders displayed cruelty and degradation towards an out-group, with these individuals illustrating anti-Jewish sentiment (44.4%) and anti-Muslim sentiment (44.4%), while one case (11.2%) involved an individual, David Bitton, who articulated both anti-Semitic and anti-Muslim sentiments, expressing antagonism towards Muslims, Jews, and individuals of Afro-Caribbean origin.²⁰²

¹⁹⁸ Ibid.

¹⁹⁹ 'Prosecutions for Antisemitic Criminal and Terrorist Acts', *Community Security Trust*, undated, available at: <https://cst.org.uk/antisemitism/prosecutions>, last visited: 10 September 2019.

²⁰⁰ Mitchell, B., and Mortimer, C., 'Mohammed Uddin: British man jailed after going to Syria to join Isis', *The Independent*, 10 February 2016, available at: <https://www.independent.co.uk/news/uk/crime/isis-syria-british-man-mohammed-uddin-a6865291.html>, last visited: 10 September 2019.

²⁰¹ 'The Counter-Terrorism Division of the Crown Prosecution Service (CPS) - Successful prosecutions since the end of 2006', *CPS*, undated, available at: <https://webarchive.nationalarchives.gov.uk/20180702132916/https://www.cps.gov.uk/counter-terrorism-division-crown-prosecution-service-cps-successful-prosecutions-end-2006>, last visited: 10 August 2019.

²⁰² 'Prosecutions for Antisemitic Criminal and Terrorist Acts', *Community Security Trust*, undated.

- There was some creation of harmful content (20%). 40% of this group were coded as 0.5, as the offender created and published content (often comments) which were interpreted as an expression of opinion. There was some sharing of content (10%), and those who created content also shared it. Mahmudul Choudhury, a teacher, was one of the individuals who fell in this category, by sharing a photo of Hitler on Facebook with his own anti-Semitic caption.²⁰³
- 15% of offenders did not express remorse. For instance, Shamim Ahmed threatened the judge at his conviction, warning him that he would “eventually come out the enemy”.²⁰⁴ Similarly, Sandeep Samra also expressed no remorse, denying that she intended to carry out acts of violence by planning to travel to Syria. Instead, she argued that she simply wanted to be in a new environment away from her Sikh family.²⁰⁵
- 5% of cases involved the individual ignoring prior warnings from technology companies, friends, members of the public, and police personnel, while also indicating no remorse. Sandeep Samra’s teachers had previously reported her to the Prevent programme, part of the UK’s counter-terrorism strategy.²⁰⁶

5.3.3. Recommendations

- In the online space, a profile-driven approach is needed for this category, as individuals tend to be secondary terrorist offenders, who are associated with main terrorist offenders. It would therefore be important to monitor behaviour that meets a threshold of indicators of online harm, to prevent the use of social media platforms to foster networks.
- A profile-driven approach should be paired with a content-driven approach, with a focus on clamping down on the individual’s glorification and justification of the use of extremist or terrorist violence, as well as online praise or support of a terrorist or extremist organisation.²⁰⁷
- Support by individuals in this group tended to focus on IS, with anti-Semitic ideology coming second. There is less nuance amongst offenders in this category, and content can be more easily linked to a proscribed group or known extremist organisation, so can be easily removed.
- Individuals in this category tended to avoid previous warnings and did not express remorse. As a result, it is possible that educative programmes or strikes to prevent individuals from using the platform are less likely to work. Instead, an approach can be taken to freeze privileges on content creation and sharing, particularly on public forums.
- Given the heavy reliance of individuals in this category on private communications, a collaborative approach between technology companies and the police will need to be taken to access case information regarding the historical behaviour of offenders, and any legal limitations on their use of the internet.²⁰⁸

²⁰³ ‘Teacher fined for posting pro-Hitler image on Facebook aimed at Jews’, *The Telegraph*, 18 February 2015, available at: <https://www.telegraph.co.uk/news/uknews/crime/11419645/Teacher-fined-for-posting-pro-Hitler-image-on-Facebook-aimed-at-Jews.html>, last visited: 10 August 2019.

²⁰⁴ ‘Isil fanatic emailed the PM’s office threatening to “wage jihad” before getting job on Crossrail’, *The Telegraph*, 14 June 2017, available at: <https://www.telegraph.co.uk/news/2017/06/14/islamic-state-fanaticemailed-pms-office-threatening-wage-jihad/>, last visited: 10 September 2019.

²⁰⁵ Eccleston, B., ‘Terror suspect tells court she wanted to move to Syria for a fresh start’, *Coventry Telegraph*, 25 January 2018, available at: <https://www.coventrytelegraph.net/news/coventry-news/terror-suspect-tells-court-wanted-14202576>, last visited: 10 September 2019.

²⁰⁶ ‘Sandeep Samra, 18, jailed for Syria terrorist acts bid’, *BBC News*, 26 January 2018, available at: <https://www.bbc.co.uk/news/uk-england-coventry-warwickshire-42835558>, last visited: 10 September 2019.

²⁰⁷ Praise, support, and recognition are already banned on the Facebook platform. For more, see: ‘Dangerous individuals and organisation’, *Facebook Community Standards*, available at: https://www.facebook.com/communitystandards/dangerous_individuals_organizations/, last visited: 17 December 2019.

²⁰⁸ This point could, of course, raise privacy concerns. There is due process through the mutual legal assistance treaty, where the proactive disclosure of data centres on real world threats or imminent harm. For more, see: ‘Mutual legal assistance treaty’, *Wikipedia*, 11 September 2019, available at: https://en.wikipedia.org/wiki/Mutual_legal_assistance_treaty, last visited: 17 December 2019.

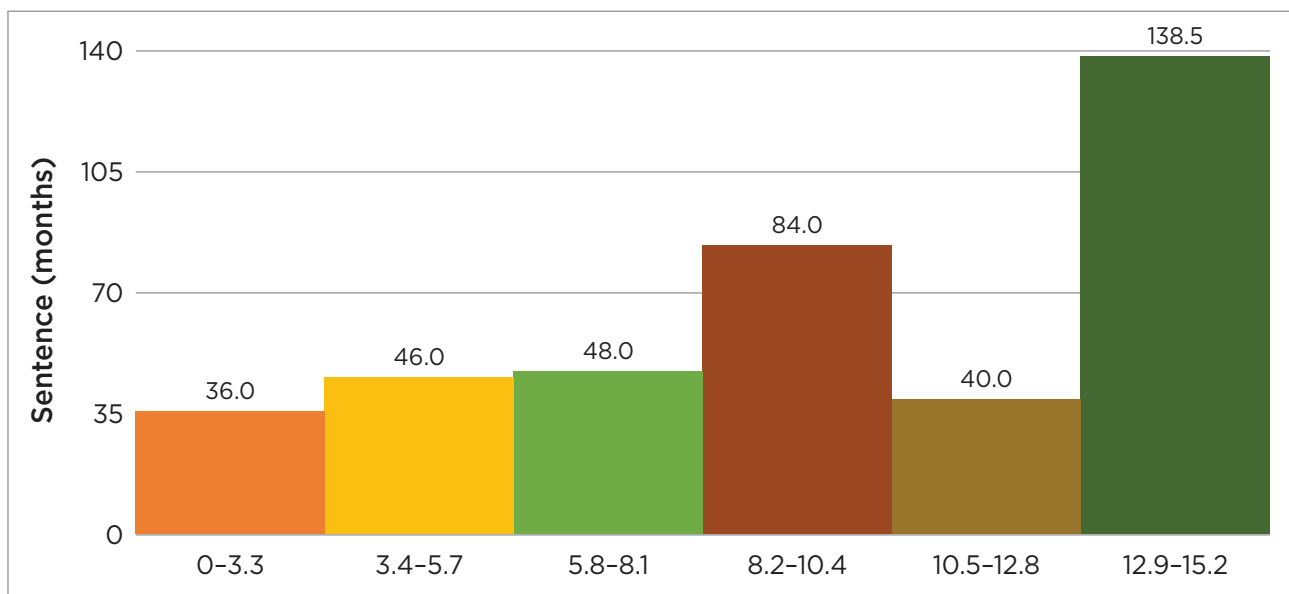
5.4. Category Ranking: 5.8-8.1

5.4.1. Analysis

The average imprisonment for the category 5.8-8.1 was 60.2 months (see Figure 10). Average imprisonment was skewed upwards by the inclusion of the profiles of Naweed Ali, Khobaib Hussain, and Mohibur Rahman, who each received 240 months, as well as Tahir Aziz and Haroon Syed, who each received 180 months. Ali, Hussain, and Rahman were graded at 7.5 on the extremism scale, Aziz was graded at 6.5, and Syed was graded at 8. Of these offenders, only Haroon Syed funded terrorist activity.

As sentences tended to be skewed upwards for the harm categories of 5.8 and above, interquartile range was used as an alternative to average imprisonment. The interquartile range imprisonment for the 5.8-8.1 category was 48 months (see Figure 13).

Figure 13: Interquartile Range Sentence in Months by Category Rankings



There continues to be a tendency for individuals in this category to commit crimes in the offline and online space. For example, Ali, Hussain, Rahman, and Aziz operated a network, and were together convicted of preparing to commit a terrorist act.²⁰⁹ Similarly, Syed was convicted for preparing to commit acts of terrorism by planning to detonate a nail bomb.²¹⁰ In contrast, offenders such as Adam Paul Wyatt were sentenced to 45 months for sharing a terrorist publication that said, “Britain must atone for its sins in Palestine”, and for posting that *jihad* was an obligation for Muslims on social media.²¹¹

As per Figure 7, 16 of the 27 cases (59.3%) were charged under the TA 2006²¹², nine (33.3%) were charged under the Public Order Act, while two cases (7.4%) were charged under the Criminal Law Act.

A comparison across categories reveals a jump in public broadcast from 40% in the 3.4-5.7 grade to 74.1% in the 5.8-8.1 grade. The most common platform utilised for offenders in the

²⁰⁹ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) - Successful prosecutions since 2016’, *Crown Prosecution Service Counter-Terrorism Division*, undated.

²¹⁰ *Ibid.*

²¹¹ ‘Salford man jailed for sharing terrorist handbooks’, *BBC News*, 19 April 2018, available at: <https://www.bbc.co.uk/news/uk-england-manchester-43816922>, last visited: 11 September 2019.

²¹² One case was a combination of a terrorism charge and a malicious communications charge, while another case was a combination of a public order charge and arson.

5.8-8.1 category was Facebook, followed by Twitter, Telegram, YouTube and WhatsApp, and Google+. The least used platforms were Skype, VK and Instagram, which were not used by any of the individuals in this category.

Out of the 27 cases in this category, 10 (37%) were associated with IS. A further 10 (37%) were not associated with an extremist or terrorist organisation, six of these 10 were convicted for anti-Muslim public rhetoric and four were convicted for anti-Semitic comments. Four cases (14.8%) were associated with AQAP, while one (3.7%) was affiliated with al-Muhajiroun (ALM), one case (3.7%) was associated with far-right group System Resistance Network, and one case (3.7%) was associated with an unidentified Islamist group.²¹³

The most common indicator in this category was 'Glorifies or justifies of the use of extremist or terrorist violence, or a past act of extremist or terrorist violence' (90.7%), followed by 'Praises or supports a proscribed terrorist organisation or extremist (individual or group)' and 'Public broadcast', both of which tied as the second most common indicator (74.1%).

18 of the 27 cases (66.7%) demonstrated an overlap between the indicators 'Glorifies or justifies the use of extremist or terrorist violence, or a past act of extremist or terrorist violence' and 'Praises or supports a proscribed terrorist organisation or extremist (individual or group)'. There was one case in which an individual praised or supported a proscribed terrorist organisation, but did not glorify or justify the use of extremist or terrorist violence, which was that of Stephen Gray, who publicly stated in his defence that he was travelling to join Ahrar al-Sham.²¹⁴ However, Gray did not praise or support the organisation in his online statements or during his trial. There were seven cases in which individuals glorified or justified the use of extremist or terrorist violence, but did not praise or support a proscribed terrorist organisation, corresponding with cases where there was no affiliation to proscribed terrorist organisations or extremist individuals or groups. These included the cases of Santinderbir Singh, John Hanson, Ian Evans, Rhodenne Chand, Andrew Emery, and Keegan Jakovlevs, who made anti-Muslim comments online. This group also included Majid Mahmood, who made anti-Semitic comments did not praise or support an individual or group.

17 of the cases in which individuals who were coded positively for 'Active member of a proscribed terrorist organisation or actively associated with an extremist individual or group' also overlapped with the indicator 'Praises or supports a proscribed terrorist organisation or extremist (individual or group)'. The three cases in which individuals praised or supported a proscribed terrorist organisation or extremist individual or group, but were not active members of a proscribed terrorist organisation were Mark Meechan, John Churchod, and Nicholas Goodwin. Meechan posted an online video of his girlfriend's dog lifting its paw to statements such as "gas the Jews" and "Sieg Heil". Churchod was charged with writing anti-Semitic comments online,²¹⁵ while Goodwin was convicted for sending an image of the Nazi flag to the Jewish mother of an acquaintance.²¹⁶ However, all three did not explicitly declare allegiance for, nor were they recognised by, far-right groups.²¹⁷

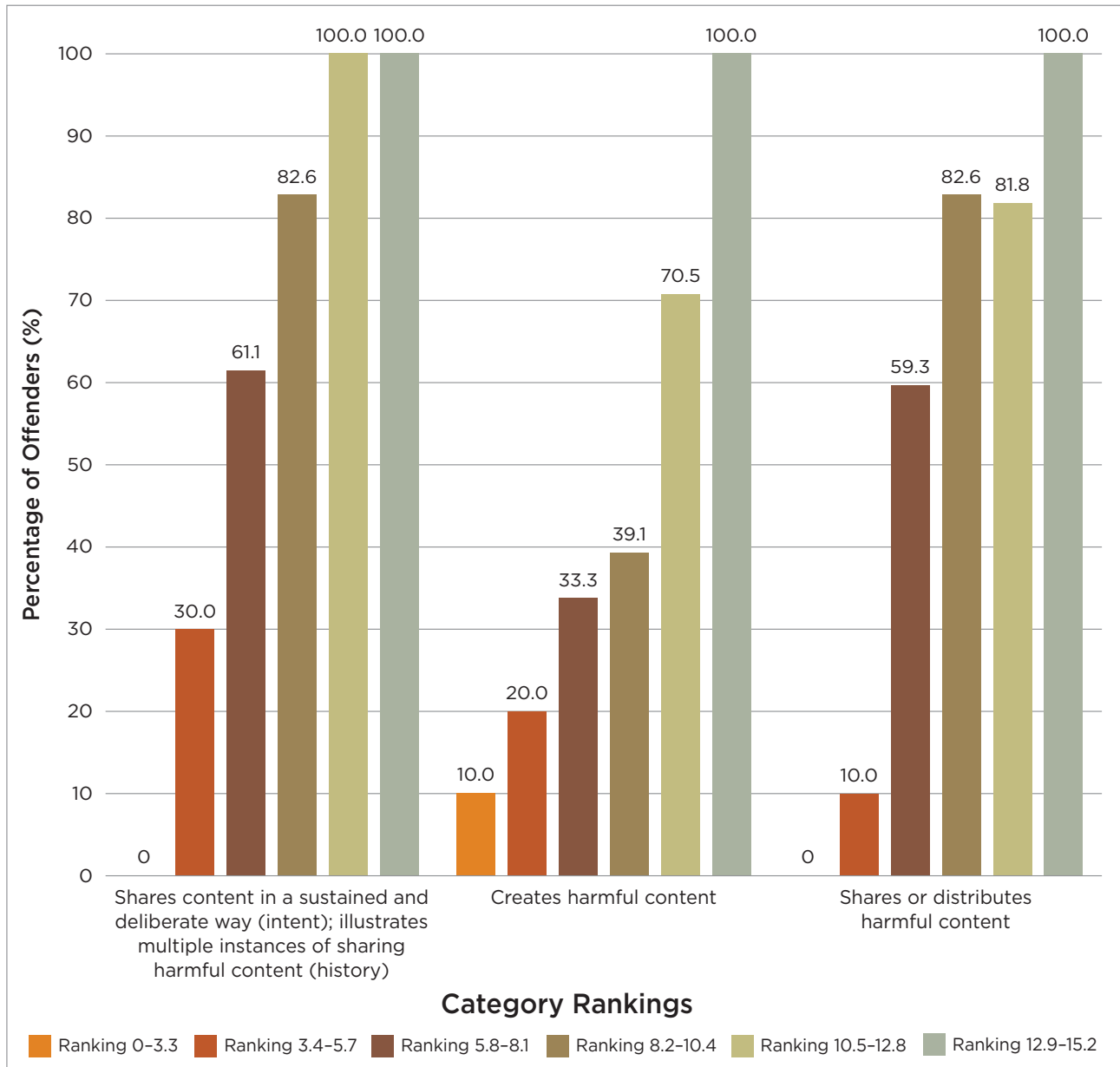
²¹³ 'Racist Newport arsonist jailed for six years', *BBC News*, 21 August 2018, available at: <https://www.bbc.co.uk/news/uk-wales-45252517>, last visited: 11 September 2019.

²¹⁴ 'RAF veteran "driven by sympathy" sentenced to 5 years for Syria terrorism offences'. *Carmelite Chambers*. Available: <http://www.carmelitechambers.co.uk/news-and-events/news/raf-veteran-driven-by-sympathy-sentenced-to-5-years-for-syria-terrorism-off>, last visited 15 November 2019.

²¹⁵ 'Prosecutions for Antisemitic Criminal and Terrorist Acts', *Community Security Trust*, undated.

²¹⁶ Beveridge, A., 'Racist thug receives six-month prison sentence after taunting Jewish victim with Nazi flag', *Daily Record*, 9 August 2015, available at: <https://www.dailyrecord.co.uk/news/scottish-news/racist-thug-receives-six-month-6221842>, last visited: 11 September 2019.

²¹⁷ Mark Meechan had been supported by Tommy Robinson, former leader of the English Defence League (EDL). However, this support was not deemed sufficient enough to warrant membership or recognition, and Meechan's attendance at the 'Day of Freedom' rally, organised by Robinson, occurred after his conviction.

Figure 14: Sharing and Creation by Category Rankings

5.4.2. Trends

- This category represents a shift in the dissemination of content that is praising and associating with an extremist or terrorist group, to a hybrid mix of influencers and ideologues. Many offenders in this category have a public broadcast element and large followings. Individuals in this category tended to focus less on terrorist funding. There is also a shift in remorse, with more people in this indicator pleading not guilty, appealing, and ignoring prior warnings.
- The average imprisonment by month in this category is 60.2 months. However, this is because five cases with particularly high sentences skew the average imprisonment length upwards. The average imprisonment length in this category is significantly reduced to 24.8 months after removing the five individuals with particularly high, anomalous sentences.
- Looking closely at the cases with particularly long sentences (60 months or higher), Naweed Ali, Khobaib Hussain, Mohibur Rahman and Tahir Aziz were collectively found guilty of preparing to commit a terrorist act, and were sentenced to 180 months each. Austin Ross, who committed acts of arson, created racist graffiti, and published links to a Hitler Youth

account on Facebook, was given a sentence of 72 months.²¹⁸ In contrast, offenders solely active in the online realm were given shorter sentences: John Hanson, who called for others to commit acts of violence against Muslims, was given a nine-month sentence. Cases in this category highlight existing challenges facing social media platforms; individuals who conduct violent actions in the offline sphere may match their offline actions with online hate speech, such as Austin Ross.²¹⁹

- There is an increase in the use of public broadcast in this category, with 20 cases (74.1%) making use of openly available social media websites. Nine cases made use of Facebook, one relied solely on Google+, one relied on Google+ and YouTube (Sagheer Hussain), three solely on Twitter, four others on Twitter and a second platform (WhatsApp, Facebook or Viber), and two solely on YouTube.
- As illustrated in Figure 14, comparing categories 3.4–5.7 and 5.8–8.1 reveals a small jump in content creation, from 20% to 33.3%. There is a far greater increase in content dissemination, from 10% of offenders to 59.3% of offenders sharing harmful content online.
- Individuals in this category tend to have a larger following than individuals in previous categories, representing a greater ability to cause harm online. Compared to 1.45% in the previous category, 35.1% of the offenders in the 5.8–8.1 category had a large following.
- There is a significant decrease in funding of violent activity, to 7.4%. There is a decrease in ‘Mentions resorting to violence with personal agency’ to 7.4%. This could represent a higher level of influence of others through public forums and content sharing, rather than intending to cause harm with personal agency.

5.4.3. Recommendations

- A purely content-based approach would overlook the danger posed by such individuals. Individuals like Austin Ross, for example, pair offline and online actions, and are therefore dissimilar from individuals classified as ‘keyboard warriors’, such as John Hanson or Keegan Jakovlevs, whose actions remain limited to the online space. Jakovlevs, in particular, was convicted for publishing a Facebook post inviting his readers to “kill every Muslim” they saw,²²⁰ but took no similar actions in the offline sphere and expressed sincere regret following his conviction.²²¹
- As such, a holistic approach needs to be developed to distinguish between individuals such as Austin Ross and Keegan Jakovlevs. Such an approach would require social media platforms to not only focus on the content posted by an individual, but also the background of a speaker, if a certain minimum threshold is crossed.²²² Focusing on banning an individual

²¹⁸ ‘Far-right supporter covered landmarks in swastikas and racist graffiti’, *Wales Online*, 21 August 2018, available at: <https://www.walesonline.co.uk/news/wales-news/far-right-supporter-covered-landmarks-15057001>, last visited: 11 September 2019.

²¹⁹ Ibid.

²²⁰ ‘Wrexham man jailed over ‘kill every Muslim’ Facebook post’, *BBC News*, 7 September 2017, available at: <https://www.bbc.co.uk/news/uk-wales-north-east-wales-41186060>, last visited: 11 September 2019.

²²¹ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) - Successful prosecutions since 2016’, *Crown Prosecution Service Counter-Terrorism Division*, undated.

²²² For example, in determining its response to harmful activity, YouTube takes into account factors including, but not limited to, the severity of the user’s actions and whether they follow a consistent pattern of harmful behaviour. This is important as it demonstrates a recognition that not all harmful actions online are identical and therefore that YouTube is in theory committed to a scaled approach of multiple possible responses to extremism rather than one in which content is simply removed. This is done through the three-strike policy. According to the policy, the first time content is removed the user responsible only receives a warning. However, successive breaches of the Guidelines are accompanied by a strike. Each strike remains in place for a 90-day period, with another infringement of the Guidelines within that period leading to a second and then third strike. Receiving a strike causes a viewer to temporarily lose the ability to upload content such as videos or playlists and make posts, first for one, and then for two, weeks. Receiving three strikes leads to the permanent removal of a channel. Any attempt to circumvent these restrictions through creating a new channel may lead to the termination of a user’s account. For more, see: ‘Community Guidelines strike basics’, *YouTube*, available at: https://support.google.com/youtube/answer/2802032?hl=en-GB&ref_topic=9387060, last visited: 11 September 2019.

from using a platform in a consistent way would likely necessitate the formation of an appeal and justification process, as well as take into account the individual's offline interactions with the legal judicial system and any prior criminal history or convictions.

- Coupling history of offences through court documents and behaviour on the online space would be especially relevant to this category, given that there is an increase in offenders who did not show remorse for their actions (14.8%), as well as those that pled not guilty (9.3%) and appealed their sentence (22.2%). Moreover, 14.8% of offenders in this category ignored warnings (see Figure 18). Individuals from this category may benefit from a mandatory educative course, completion of which is required to re-establish their use of social media platforms.

5.5. Category Ranking: 8.2-10.4

5.5.1. Analysis

The average sentencing in this category was 79.8 months, the highest across all categories (see Figure 10). Average sentencing was heavily skewed by the inclusion of Mohammed Rehman, who received a sentence of 324 months for plotting an attack on the tenth anniversary of the 7/7 London bombings.²²³ Excluding Rehman yields an average sentence of 68.2 months; the interquartile range of sentencing for this category was 84 months (see Figure 13).

The most common legislation employed for offenders in this category was the Terrorism Acts 2000 and 2006 (78.3%). 17.4% of offenders were convicted under the Public Order Act.

As per Figure 12, 100% of individuals in this category glorified or justified the use of extremist or terrorist violence or a past act of extremist or terrorist violence. The majority of individuals (78.3%) identified with IS, or with IS and another radical Islamist organisation such as AQAP. One out of 23 cases (4.3%) was affiliated with AQAP and Al-Muhajiroun, a radical Islamist group based in the UK. Four out of 23 (17.4%) demonstrated no allegiance to any proscribed militant organisation, but had exhibited anti-Muslim and/or anti-Semitic rhetoric online.

This category represented a shift on the part of offenders from focusing less on offline acts, and more on online incitement. Offenders in this category tended to show more intent to incite violent acts through the dissemination of harmful content. A category comparison illustrates an increase in persistent sharing and distribution of harmful content, as compared to lower categories (see Figure 14).

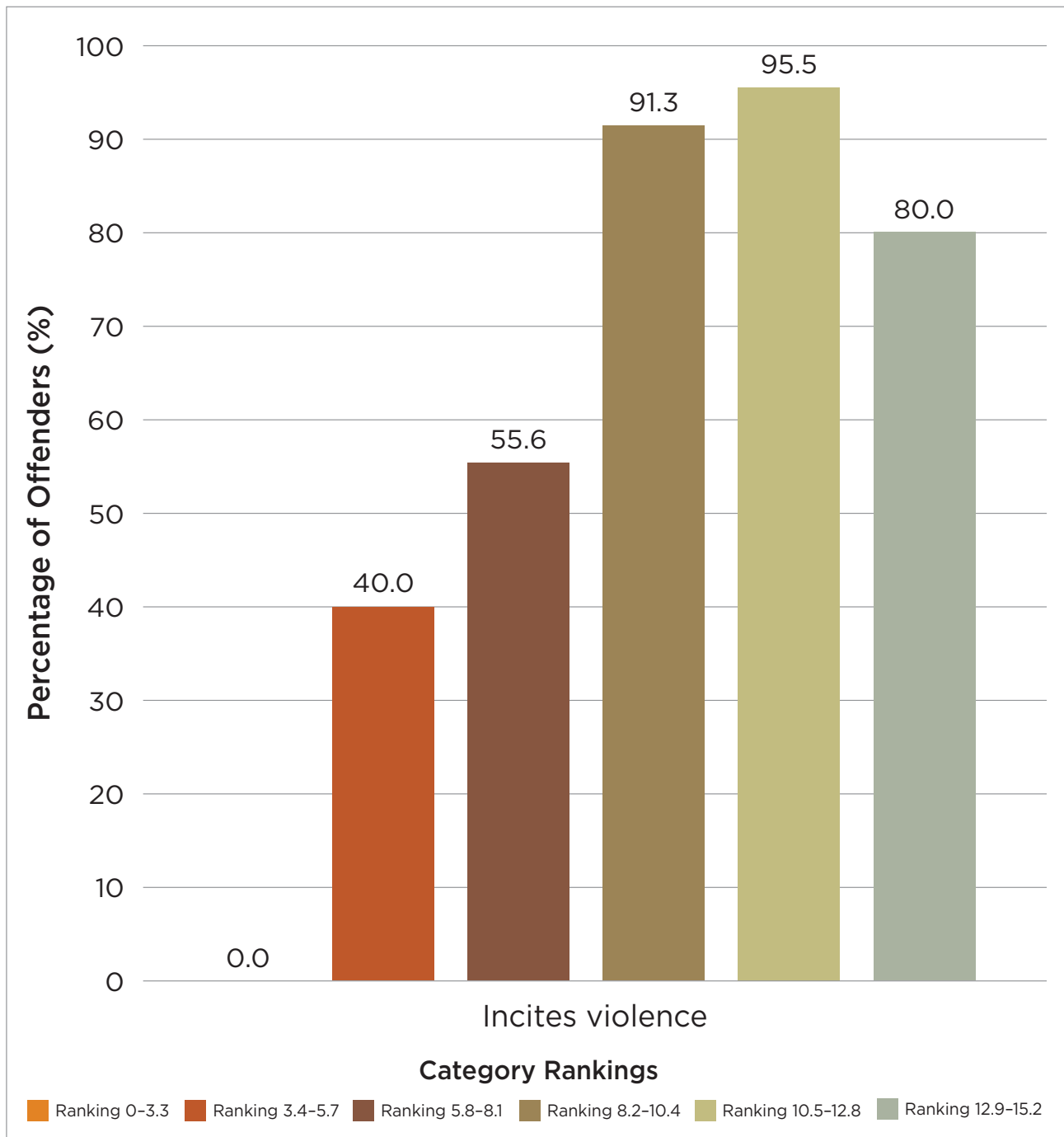
As illustrated by Figure 15, this category witnessed a leap in offenders who incited violence – from 55.6% in the 5.8-8.1 category, to 91.3% in the 8.2-10.4 category. Key offenders exemplifying this trend and included in this category were Zana Abbas Sulieman, who relied on three Facebook accounts to share IS propaganda videos and a magazine urging individuals to target diplomats and embassies.²²⁴

A significant number of cases (87%) praised or supported a proscribed terrorist organisation or extremist individual or group. All individuals who praised or supported a proscribed terrorist organisation also glorified or justified the use of extremist or terrorist violence, except for Peter John Tovey, Shane Fletcher, and Andrew Littlefair. The latter three cases did not praise or support a proscribed terrorist organisation by publicly expressing support, or publishing violent or non-violent imagery for a proscribed terrorist organisation and/or an extremist organisation or group. Rather, all three made anti-Muslim comments and/or anti-Semitic

²²³ 'Jailed "Silent Bomber" had explosives recipe in cell', *BBC News*, 24 February 2017, available at: <https://www.bbc.co.uk/news/uk-england-london-39076416>, last visited: 12 September 2019.

²²⁴ 'The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016', *CPS*, undated.

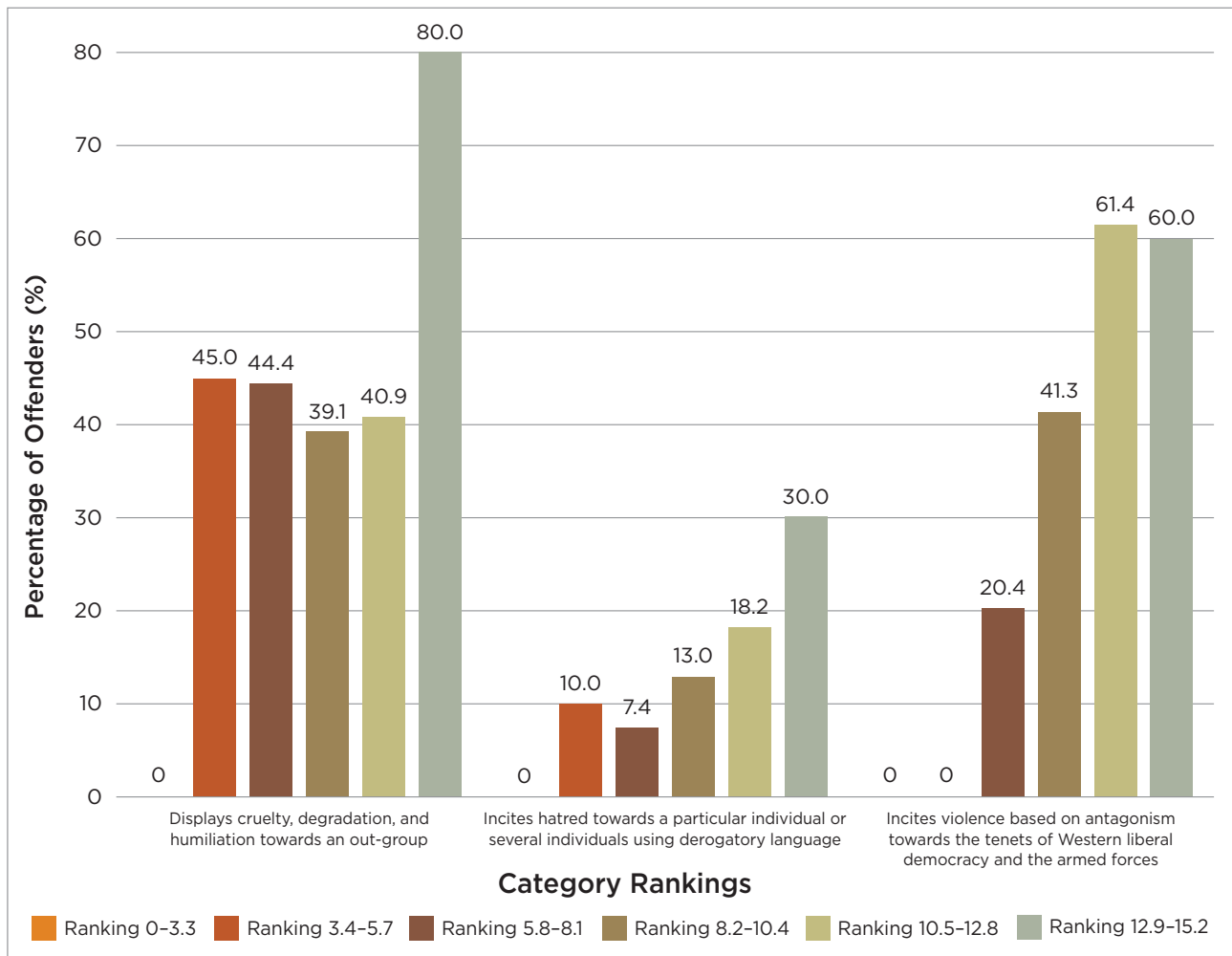
Figure 15: Incitement of Violence by Category Rankings



comments, with Fletcher in particular being convicted of plotting a massacre in his hometown of Workington.²²⁵

There was a large crossover between individuals who praised or supported a proscribed terrorist organisation (87%) and individuals who were fully or partially active members of a proscribed terrorist organisation (56.5%) (see Figure 17). Tovey, Fletcher, and Littlefair (who did not praise or support a proscribed terrorist group) were also not active members of such groups. This was understandable given that they did not declare allegiance for, nor were they recognised by, terrorist organisations.

²²⁵ 'Shane Fletcher convicted of Workington massacre plot', *BBC News*, 24 January 2019, available at: <https://www.bbc.co.uk/news/uk-england-cumbria-46987390>, last visited: 12 September 2019.

Figure 16: Incitement and Cruelty by Category Rankings

Nonetheless, the case of Muhammad Hamza Siddiq saw an instance of someone who was not an active member of a proscribed terrorist organisation, but publicly expressed support for the group. Siddiq had expressed support for IS, but he did not declare allegiance for, nor was he recognised by, the group. Siddiq explicitly stated that he “[did] not think the attack in Manchester [a reference to the Manchester arena bombing on 22 May 2017] was the best choice of action to take” but described it as an “inevitable consequence of British foreign and domestic policy towards Muslims”.²²⁶

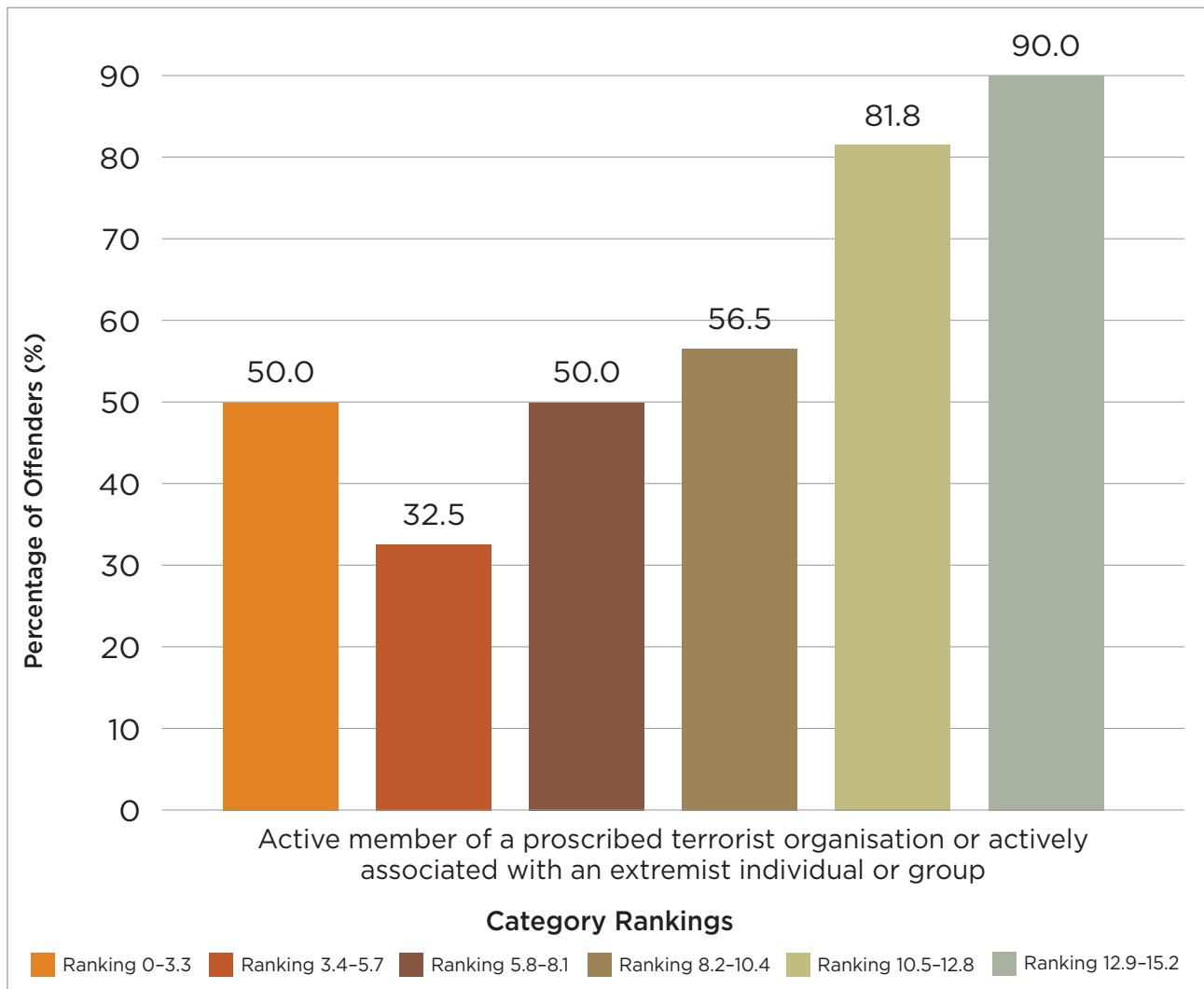
5.5.2. Trends

- There is a large increase in incitement of violence from 55.6% for the 5.8-8.1 category to 91.3% for the 8.2-10.4 category. The two individuals in the 8.2-10.4 category who did not incite violence were Mourad Mosdefaoui and Mohammed Rehman. Both shared similarities in that they made reference to the utilisation of armed violence, but in ways that avoided encouraging individuals to undertake violent acts. For instance, Mosdefaoui published posts asking individuals to pray for him that he may be used to “defend his religion and the honour of the virtuous girls in Iraq and Syria”,²²⁷ but he did not incite individuals to do the same. Similarly, Mohammed Rehman published a tweet that asked online users to

²²⁶ ‘Muslim convert screams “Allahu Akbar” as he is jailed for encouraging terrorism on Facebook’, *Nottinghamshire Live*, 22 February 2019, available at: <https://www.nottinghampost.com/news/local-news/muslim-convert-screams-allahu-akbar-2573997>, last visited: 12 September 2019.

²²⁷ Stuart, H. ‘Islamist Terrorism: Analysis of offences and attacks in the UK (1998-2015)’, *The Henry Jackson Society*, 2017, p.913.

Figure 17: Membership by Category Rankings



pick between two locations for a bomb attack which he intended to carry out, but did not directly incite individuals to commit the attack.²²⁸

- The most frequently used platform by offenders in this category was Facebook, followed by Twitter; WhatsApp and Telegram; Instagram, YouTube and PalTalk; and finally, VK.
- As illustrated by Figure 16, there was a rise in offenders inciting hatred towards a particular group or individual or several individuals using discriminatory harmful language, from 7.4% in the 5.8-8.1 category to 13% in the 8.2-10.4 category. The three individuals identified in this category were Mourad Mosdefaoui, Ahmed Hussain, and Saer Hussain. Hussain posted a video depicting leaders in Iraq with former US Presidents George Bush and Barack Obama, and US officers; the video subsequently showed multiple men being killed.²²⁹
- This category also represented an increase in offenders inciting violence based on antagonism towards the tenets of Western liberal democracy and the armed forces, from 20.4% in the 5.8-8.1 category to 41.3% in the 8.2-10.4 category (see Figure 16). Cases in this category included individuals such as Lloyd Gunton, who wrote a note of his intentions to

²²⁸ “‘Silent Bomber’ Mohammed Rehman and wife Sana Ahmed Khan sentenced to life in prison for London terror plot”, *The Independent*, 30 December 2015, available at: <https://www.independent.co.uk/news/uk/crime/silent-bomber-mohammed-rehman-and-wife-sana-ahmed-khan-sentenced-to-life-in-prison-for-london-terror-a6790741.html>, last visited: 12 September 2019.

²²⁹ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016’, *CPS*, undated.

“run down non-believers with a car”²³⁰ and who had researched carrying out attacks on targets including Cardiff Castle.²³¹

- There were no cases of individuals funding violent activity in this category.
- There was also a rise in offenders mentioning resorting to violence with personal agency, from 7.4% in the 5.8–8.1 category to 39.1% in the 8.2–10.4 category. Individuals who were coded positively for this indicator included Zakariya Ashiq who made explicit online threats involving a personal element to engage in a terrorist act. Ashiq noted that he would do “Istishhad (martyrdom) against any ... all these people” any chance he got.²³² Similarly, the note Lloyd Gunton left behind declared that Gunton was a “soldier of the Islamic State” and detailed his planned attack on Cardiff.²³³
- 82.6% of offenders in this category shared content in a sustained and deliberate way, showing significant intent. This was an increase from the previous category, from 61.1% in the 5.8–8.1 category to 82.6% in the 8.2–10.4 category.
- There was a significant rise in the percentage of individuals engaged in the sharing or distribution of harmful content, from 59.3% in the 5.8–8.1 category to 82.6% in the 8.2–10.4 category. Of the individuals who shared content in a sustained and deliberate way, four out of the 19 offenders shared content solely on Facebook, five offenders shared content on Facebook and another social media platform (Telegram, VK, YouTube, Gmail, Twitter), two offenders shared content solely on Twitter, and three offenders shared content on Twitter and another social media platform (Facebook, WhatsApp and YouTube), with one overlap involving Facebook and Twitter. One offender shared content solely on Instagram, while one other offender shared content on Instagram and another platform (WhatsApp). Lastly, two offenders shared content on PalTalk, one on Tor and PGP, and one on Blackberry Messenger.
- There remained a strong public broadcast element in this category, with 69.6% of offenders relying on public broadcast.
- This category witnessed some creation of harmful content (39.1%). Nine out of the 13 cases involved content creation related to IS, while one case constituted content creation related to IS and AQAP. One out of the 13 cases was related to anti-Muslim sentiment, but was unaffiliated with a particular group, and two others were related to anti-Semitic and anti-Muslim sentiment, but similarly unaffiliated with a particular group.
- A category comparison illustrates an increase in the lack of remorse indicated, from 14.8% of offenders who showed no remorse in the 5.8–8.1 category to 30.4% in the 8.2–10.4 category. Similarly, there was a rise in the percentage of offenders who pleaded not guilty from 9.3% in the 5.8–8.1 category, to 21.7% in the 8.2–10.4 category.
- 10 out of 23 cases (43.5%), or almost half the offenders in this category, ignored warnings from technology companies, friends, members of the public, or the police.
- 7 out of the 10 cases who ignored warnings from technology companies, friends, members of the public or the police also shared content in a sustained and deliberate way with intent. This included Sean Creighton who made numerous posts every day and had been blocked

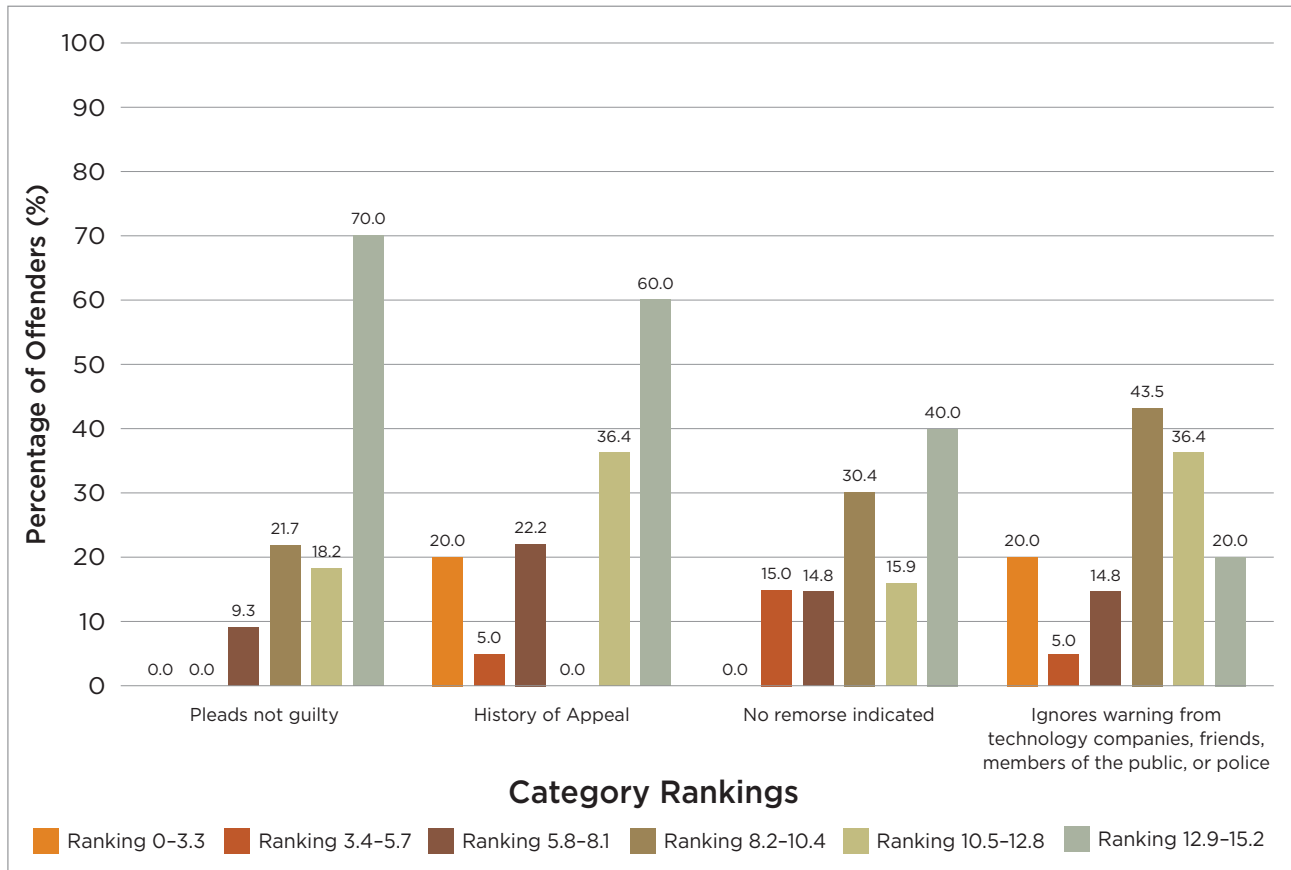
²³⁰ Morris, S., ‘Teenager given life sentence for planned Justin Bieber gig attack’, *The Guardian*, 2 March 2018, available at: <https://www.theguardian.com/uk-news/2018/mar/02/teenager-lloyd-gunton-jailed-cardiff-pop-concert-terrorist-attack-plot>, last visited: 12 September 2019.

²³¹ ‘Schoolboy given life sentence for plotting Westminster style terror attack in Cardiff’, *Crown Prosecution Service Counter-Terrorism Division*, 2 March 2018, available at: <https://www.cps.gov.uk/cymruwales/news/schoolboy-given-life-sentence-plotting-westminster-style-terror-attack-cardiff>, last visited: 12 September 2019.

²³² ‘Zakariya Ashiq jailed for six years for trying to join the Islamic State’, *BBC News*, 27 May 2015, available at: <https://www.bbc.co.uk/news/uk-england-coventry-warwickshire-32900923>, last visited: 12 September 2019.

²³³ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016’, *CPS*, undated.

Figure 18: Appeals Process by Category Rankings



by Facebook more than 300 times before he switched to VK.²³⁴ Similarly, Sabbir Miah disseminated Islamist extremist propaganda on Facebook and Instagram, but continued posting videos promoting IS martyrdom videos on Instagram and Whatsapp, even after he was released on bail.²³⁵ Abdulrahman Alcharbati was convicted of making 70 posts on Facebook in support of IS.²³⁶

- The three cases in which individuals ignored warnings from technology companies and others, but did not share content in a sustained and deliberate way with intent, were Shane Fletcher, Jabed Hussain, and Zakariya Ashiq. All three individuals relied on private rather than public channels and their actions were based in the offline rather than online domain. Fletcher, for instance, had been referred to the Prevent programme for expressing right-wing anti-Semitic views following his previous conviction for arson. However, he was convicted for planning an attack on a local sporting event and relied largely on private Facebook messages, rather than public channels, to encourage a friend to join in the attack.²³⁷ Similarly, Jabed Hussain was planning to travel to Syria to join IS, and was primarily using the social media platform Surespot to communicate and plan logistics.²³⁸

²³⁴ Robinson, M., ‘Racist bigot who called Hitler “God” and had a swastika tattoo on his chest is jailed for five years for having a “bomb recipe” on his laptop and posting hundreds of extremist messages online’, *Daily Mail*, 23 February 2017, available at: <https://www.dailymail.co.uk/news/article-4253174/Racist-bigot-swastika-tattoo-chest-jailed.html>, last visited: 13 September 2019.

²³⁵ ‘Muslim extremist jailed for spreading ISIS propaganda on Facebook’, *Court News UK*, 19 June 2017, available at: <https://courtnewsuk.co.uk/muslim-extremist-jailed-spreading-isis-propaganda-facebook/>, last visited: 13 September 2019.

²³⁶ ‘Abdulrahman Alcharbati guilty of IS Facebook posts’, *BBC News*, 19 October 2018, available at: <https://www.bbc.co.uk/news/uk-england-tyne-45917767>, last visited: 13 September 2019.

²³⁷ ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since 2016’, *CPS*, undated.

²³⁸ *Ibid.*

5.5.3. Recommendations

- Cases included within this category tend to be more dangerous (as evidenced by a jump in sentencing for offenders in this category), due primarily to the risk of potential violence occurring in the offline space linked to incitement in the online space. Technology companies should therefore focus on monitoring incitement and intent, given the rise in offenders mentioning resorting to violence with personal agency from 7.4% in the 5.8–8.1 category to 39.1% in the 8.2–10.4 category.
- Individuals in this category tended to share or distribute harmful content, and demonstrated repeated attempts to share such harmful content, scoring high on the indicator for sharing content in a sustained and deliberate way. Individuals also tended to continue to post, in spite of warnings, or navigated to an alternative platform such as VK. Social media companies could consider making stronger warnings for individuals within this category, and also increasing the duration for which individuals are prevented from posting, concomitant with the number of warnings issued to individuals.
- There is a big jump in attempting to radicalise others in this category as compared to previous categories, with 47.8% of offenders attempting to influence their audience in a harmful way. Potential inhibitors which could be examined by social media companies would be not allowing users in this category to communicate publicly and creating inhibitors on advertising and publicly live streaming contingent on warnings issued.²³⁹
- As the category rankings increase, the role between the police and technology companies are reversed. Once a number of indicators are deemed positive in the online space, it will be important for technology companies to inform the police of cases on their platform where users are exhibiting harmful behaviour, to reduce the risk of violence occurring offline.

5.6. Category Ranking: 10.5-12.8

5.6.1. Analysis

The average imprisonment for this category was 52.2 months (see Figure 10). The inclusion of the case of Husnain Rashid, who was sentenced to 300 months, was a significant outlier skewing the database upward; the interquartile range of sentencing for this category was 40 months (see Figure 13). Rashid had a prolific social media presence online, and called for followers to engage in attacks, including the targeting of Prince George at his school and attacks on football stadiums.²⁴⁰

Individuals in this category tended to be sentenced primarily under the Terrorism Acts 2000 and 2006, as opposed to the Public Order Act. 18 cases (81.8%) were charged under the Terrorism Acts 2000 or 2006, one case was charged under the Communications Act 2003, one case was charged under the Public Order Act 1986, one other was charged under the Public Orders Act 1986 and the Malicious Communications Act and one remaining case was charged under the Criminal Damage Act 1971 and the Public Order Act 1986 (see Figure 7).

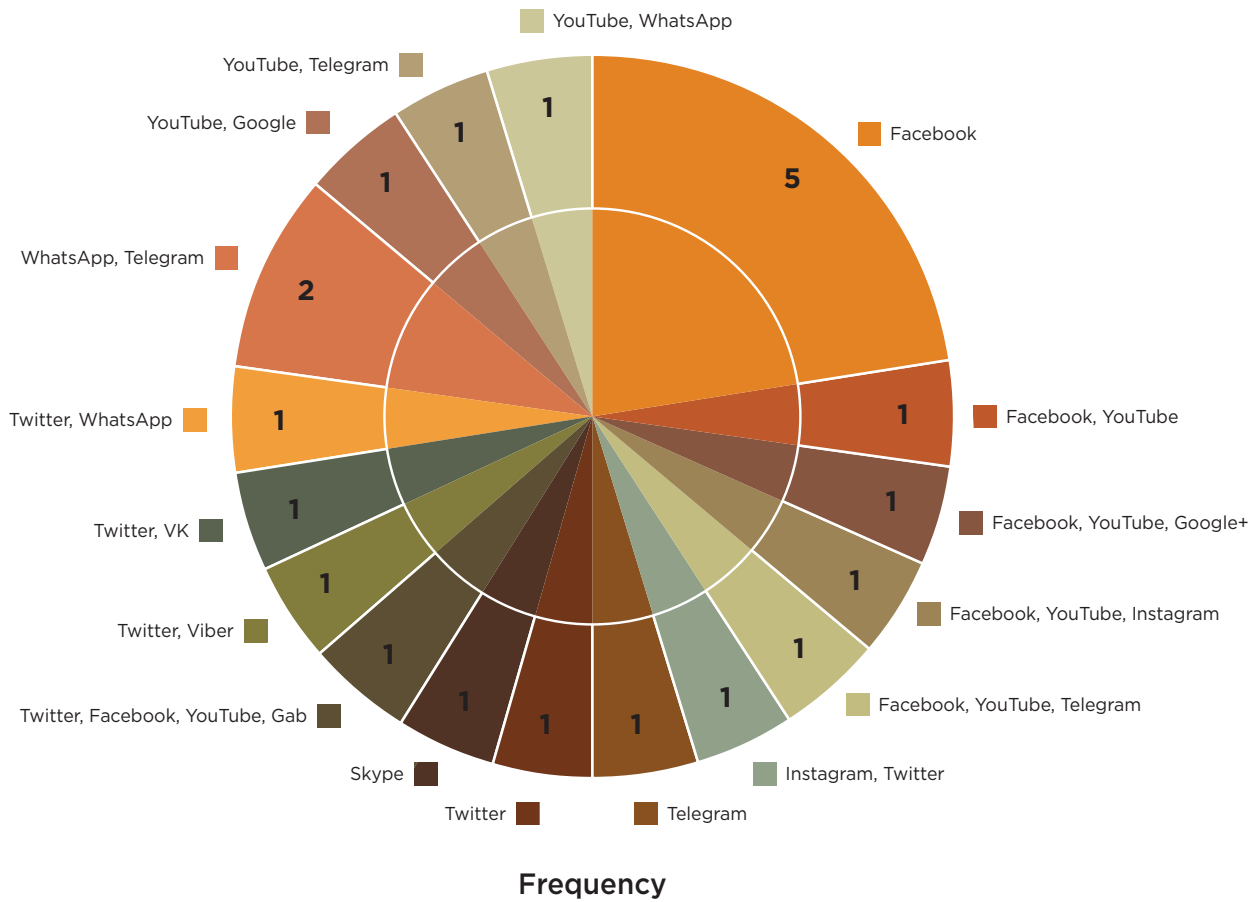
5.6.2. Trends

- This category witnessed an increase in the sharing of content. All of the cases in this category shared content in a sustained or deliberate way (illustrating intent). Offenders

²³⁹ See, for example: 'Protecting Facebook Live from abuse and investing in manipulated media research', *Facebook*, 14 May 2019, available at: <https://about.fb.com/news/2019/05/protecting-live-from-abuse/>, last visited: 17 December 2019.

²⁴⁰ 'Islamic State supporter Husnain Rashid jailed for life over Prince George plot', *Sky News*, 13 July 2018, available at: <https://news.sky.com/story/islamic-state-supporter-husnain-rashid-jailed-for-life-over-prince-george-plot-11435980>, last visited: 13 September 2019.

Figure 19: Use of Social Media Platforms in Category Ranking 10.5-12.8



used multiple platforms, as illustrated by Figure 19, with the most common being Facebook (five cases) or Facebook with a second social media platform such as YouTube, Google+, Instagram, or Telegram (five cases). The second most common social media platform used was YouTube alongside other social media platforms (eight cases).

- 21 of the 22 cases (95.5%) involved an incitement of violence.
- 11 of the 22 cases (47.5%) had a large following. This is the second highest score across all categories. The only group with a larger percentage of individuals with a large following is the last, and most harmful, category of 12.9-15.2, with 79.6% of cases having a large following.
- 86.4% of offenders in this category used public broadcast.
- As illustrated by Figure 17, this category had the second highest percentage of offenders who were active members of a proscribed terrorist organisation or actively associated with an extremist individual or group (81.8%). This is an increase from 56.5% in the 8.2-10.4 category, but still lower than the 90% of offenders in the 12.9-15.2 category.
- 20 out of the 22 cases in this category affiliated with a proscribed or dangerous organisation. The main groups in this category were IS (16 of 22 cases or 72.7%). There was one case affiliated respectively with AQAP, the EDL, and National Action, as well as one linked to unspecified radical Islamist groups in Syria.
- There was an increase in content creation, from 39.1% in the 10.5-12.8 category to 70.5% in the 10.5-12.8 category.
- However, there was a reduction in the percentage of offenders in this category ignoring warnings from technology companies, friends or members of the public, or police, decreasing from 43.5% in the 8.2-10.4 category to 36.4% in the 10.5-12.8 category. Similarly,

there is a decrease in the percentage of individuals who did not indicate remorse, from 30.4% in the 8.2–10.4 category to 15.9% in the 10.5–12.8 category. In addition, there was a decrease in the percentage of individuals in this category pleading not guilty, from 21.7% in the 8.2–10.4 category to 18.2% in the 10.5–12.8 category.

5.6.3. Recommendations

- Content sharing remains important for offenders in this category. This may be correlated with more charismatic leadership: one example of an offender in this category is Boy X (also known as RXG), who sent more than 3,000 messages to 18-year old Svedet Besim in Melbourne, inciting Besim to conduct an attack during the 2015 Anzac Day parade in Melbourne, Australia.²⁴¹ X was also in a romantic relationship with another offender, Y, and had made plans to travel to Syria together with them.²⁴² Offenders in this category are more focused on sharing content in a sustained way, with intent and deliberation. They have less interaction with companies or the public.
- Social media companies should increasingly look into a holistic content- and speaker-driven approach for offenders in this category, to capture elements of an individual's influence on followers on social media platforms that may not yet be detected in the real world or by the police. This would include examining the violence inherent in posts (95.5%) and updating of policies on sharing with intent (100%). This could include implementing penalties inhibiting individuals from actively seeking and addressing an audience, such as a prohibition on live streaming, hosting events on Facebook, advertising, the blue tick of verification, and sending messages to large groups.
- As the above cases indicate, individuals in this category are more likely to show remorse or change views. As a result, they could benefit from prior warnings (only eight of the 22 cases received prior warnings) and/or a mandatory educative course, completion of which is required to re-establish their use of social media platforms.

5.7. Category Ranking: 12.9–15.2

5.7.1. Analysis

The average sentence for this category is 76.6 months (see Figure 10). Average sentencing is skewed upwards with the inclusion of the case of Brustholm Ziamani, who received a sentence of 264 months. Ziamani was convicted of planning a terrorist attack targeting British soldiers, and had shared comments on his Facebook account stating that he was “willing to die in the cause of Allah”.²⁴³ The average sentence for this category is significantly reduced from 76.6 months to 29.8 months after the removal of Ziamani's case; the interquartile range for sentencing in this category was 138.5 months (see Figure 13).

Two out of five of the cases in this category were charged under TA 2006, while two others were charged under the Public Offences Act 1986. The fifth and final case was charged under the Communications Act 2003.

5.7.2. Trends

- Every case included in this category glorified or justified using extremist or terrorist violence, or a past act of extremist or terrorist violence.

²⁴¹ ‘Anzac Day plot: Terror teen ‘faces attack if identified’, *BBC News*, 20 November 2018, available at: <https://www.bbc.co.uk/news/uk-england-lancashire-46277360>, last visited: 13 September 2019.

²⁴² ‘The Counter-Terrorism Division of the Crown Prosecution Service (CPS) – Successful prosecutions since the end of 2006’, *Crown Prosecution Service Counter-Terrorism Division*, undated.

²⁴³ ‘Teen gets 22 Years for Plot to Behead Soldier’, *Sky News*, 20 March 2015, available at: <https://news.sky.com/story/teen-gets-22-years-for-plot-to-behead-soldier-10366930>, last visited: 13 September 2019.

- Every individual in this category praised or supported a proscribed terrorist organisation or extremist (individual or group). Out of the five cases, three were affiliated with IS, one with the far-right group London Forum, and one with proscribed organisation National Action.
- Every case in this category shared content in a sustained and deliberate way, showing intent. The most commonly used platform was YouTube, followed by Facebook, with Instagram, WhatsApp and Google+ collectively coming third.
- All cases included elements of public broadcast. Three of the five cases involved YouTube and a second social media platform (Google+, own website, or Facebook). One case involved Instagram and one other case involved Facebook and WhatsApp.
- Every offender in this category created harmful content. Alison Chabloz, included in this category and in the qualitative case study earlier in the report, wrote and performed anti-Semitic songs that denied the Holocaust at a London Forum meeting; she additionally shared her songs on YouTube and her own website.²⁴⁴
- Every case in this category shared or distributed harmful content.
- 90% of offenders in this category were active members of a proscribed terrorism organisation, or actively associated with an extremist individual or group.
- 80% of offenders incited violence, with 80% displaying cruelty and degradation towards an outgroup. Akeem Samuels, for example, shared images on Instagram, including one of a forearm being cut with a knife and the caption, “Shi’ism - cut it out before it spreads”.²⁴⁵ Separately, Lawrence Burns shared harmful and discriminatory comments against the Jewish community and individuals of Afro-Caribbean origin.²⁴⁶
- 60% of offenders incited violence towards the West. Akeem Samuels uploaded a post containing the silhouettes of four people holding machine guns with the message: “Fight those who do not believe in Allah” and mocked the Orlando shooting massacre.²⁴⁷
- 60% of cases mentioned resorting to violence with personal agency.
- Many of the cases involved appeals against their convictions: four of the five cases (80%) involved partial or full non-guilty pleas, three of five cases (60%) involved appealing existing sentences, and two of the five cases (40%) involved a lack of remorse.
- Only one of the five cases involved the individual ignoring prior warnings from technology companies, friends, members of the public, or police. Brustholm Ziamani had been previously arrested by police personnel for a letter which he wrote about attacking British soldiers.²⁴⁸ He had additionally rejected offers of help from Prevent, part of the UK’s counterterrorism programme.²⁴⁹
- None of the cases involved funding violent activity.

²⁴⁴ ‘Prosecutions for Antisemitic Criminal and Terrorist Acts’, *Community Security Trust*, undated.

²⁴⁵ ‘ISIS supporter posed as Osama bin Laden, posted images of decapitated bodies and mocked victims of Orlando massacre’, *Mirror*, 29 April 2017, available at: <https://www.mirror.co.uk/news/isis-supporter-posed-osama-bin-10320084>, last visited: 13 September 2019.

²⁴⁶ ‘Cambridge extremist who admired Hitler and wanted to “hang the black race” found guilty of stirring up racial hatred’, *Cambridge News*, 15 December 2016, available at: <https://www.cambridge-news.co.uk/news/cambridge-news/cambridge-extremist-who-admired-hitler-12327721>, last visited: 13 September 2019.

²⁴⁷ Haddad, T., ‘London Isis propagandist Akeem Samuels jailed for four years’, *International Business Times*, 29 April 2017, available at: <https://www.ibtimes.co.uk/london-isis-propagandist-akeem-samuels-jailed-four-years-1619292>, last visited: 13 September 2019. See also: https://terrorismlegislationreviewer.independent.gov.uk/wp-content/uploads/2018/10/The_Terrorism_Acts_in_2017.pdf.

²⁴⁸ Walker, P., ‘Teenager jailed for 22 years for plotting to copy Lee Rigby beheading’, *The Guardian*, 20 March 2015, available at: <https://www.theguardian.com/uk-news/2015/mar/20/brustholm-ziamani-jailed-plot-behead-soldier-lee-rigby>, last visited: 13 September 2019.

²⁴⁹ Israel, S., ‘London teenager Ziamani guilty of plot to behead soldier’, *4 News*, 19 February 2015, available at: <https://www.channel4.com/news/brustholm-ziamani-guilty-behead-soldier-london-teenager>, last visited: 13 September 2019.

5.7.3. Recommendations

- Individuals in this category should be banned, due to their extensive capacity for harm in the online and offline domain. This is evidenced by their large followings, their creation and sharing of harmful content, and their ability to share material on public platforms in a sustained and impactful way.²⁵⁰
- Although individuals in this category are less likely to show remorse or change views nonetheless, they could benefit from prior warnings (only one of the five cases had received warnings).
- There is a need to de-platform or de-legitimise individuals in this category, as they represent the highest level of harm with intent.
- It would be important to action strikes or removal of content against individuals in this category quickly, given that their propensity to harm and influence is higher than those of lower categories.

²⁵⁰ It could be argued that some indicators can, and should, in themselves lead to a ban. However, the cumulative approach put forward in this report marries different levels of harm with different restrictions, as opposed to a simple 'ban or no-ban' option. As such, a certain threshold of harm would need to be met before the punitive action of a ban is taken. Moreover, the effectiveness of a ban can in and of itself be questioned, given that organisations and individuals are able to operate under new names to avoid detection, as has been the case offline. A multitude of different restrictions would allow people to communicate within boundaries, rather than not at all.

6. Broader Recommendations and Conclusions

The recommendations contained within this report argue for a degree of further regulation in the online domain. In order to find a balance between protection, toleration, and freedom when regulating extremist content, multiple parties (technology companies, government departments, and the police) should use a tiered-approach to dismantling online privileges that examines the three aspects of speech that give it the capacity to harm: the speech, the speaker, and the space.

Rather than the binary action of being permitted or unpermitted to preach hateful content, the report has studied the applicability of a 'tiered' approach to dismantling the privileges of those who propagate extremist views. The 'tiered' approach can, and should, be linked and correlated with decisions made against extremist speakers offline, including but not limited to legal judgements, using the case studies and examples developed in the data analysis sections of this report. The resulting actions may include, for example, a removal of public accessibility based on popularity of the user, context of the comment, history of the offender, and willingness to engage with educative and/ or rehabilitative aspects of the platform in question. Timescales of privilege removal and the ability for users to appeal and renege views in order to be allowed full privileges should also be developed, comparable to a legal action and appeals processes offline. This would help ensure that content that legitimately criticises government policies or religion, or uses art and satire to do so, is not caught up in a wide net of 'extremist' views.

Key to understanding when content, or individuals, are seen as extremist online is the idea of tolerability. This report argues that the use of several indicators, some of which occur offline and link to an individual's history (but can be easily accessible on the open source, such as court documents) lend credibility to the idea that speech is more harmful when it is propagated with intent, against a particular audience, or by an influential speaker. Therefore, technology companies should couple a threshold of key online indicators being met with offline material to flag individuals who may be exploiting their platforms. This is certainly not an easy task: similar initiatives to list extremist speakers in the offline space in the UK have met with controversy.

Nonetheless, proposals put forward by the Commission for Countering Extremism (CCE) argue that organisations which promote hateful extremist agendas – aimed at the destruction or restriction of the rights of others and wider fundamental freedoms – should not be tolerated or endorsed.²⁵¹ This report would take this one step further and apply a lack of toleration towards individuals who are extremist as well: some of whom hold positions of power and authority and are able to influence audiences with their views online. Early action and coordination between the online and offline space can help mitigate a speaker or organisation from further gaining influence. Better transparency around why certain organisations and individuals are deemed 'extremist' will be necessary, with patterns of behaviour being shared in case studies by technology companies in annual reports and in any summits where multiple stakeholders meet to discuss the viability of regulating harmful material online. Unlike the offline space, where banning or naming an organisation as extremist may come with a unique set of legal challenges, technology companies are able to limit a user's ability to take advantage of their products due to a violation of terms and conditions. This must be done in a consistent and transparent way, with coordination across bodies in government spaces and with other technology companies.

²⁵¹ 'Challenging Hateful Extremism', Commission for Countering Extremism, October 2019, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/836538/Challenging_Hateful_Extremism_report.pdf, p. 122, last visited: 10 October 2019.

Moreover, banning an organisation or individual may not always be effective in the long term, as users find new and innovative ways to mitigate bans or disseminate ideas (including, but not limited to, rebranding or splintering, ‘dog whistles’, flipped imagery, or using other individuals to disseminate the same ideas). A unique set of tools based on an extremist spectrum of harm would facilitate the removal of different privileges to limit a speaker’s influence, while still allowing for the justification that terms of service have been violated.

As such, the report puts forward the following recommendations:

1. *Across multiple stakeholders*

- A consistent framework of ‘online extremist harm’ should be used to assess and flag patterns of behaviour, focusing on both violent and non-violent (harmful) extremism.²⁵²
- Once an initial threshold of online warning signs is met (measured by three or more online indicators being flagged) technology companies should work closely with police or consult open source court material to understand the background of those convicted of terrorism or extremist offences.
- Rather than the binary or non-binary action of banning an individual or organisation for extremist behaviour, or removing extremist content, technology companies and courts can consider tailored approaches based on similarities of those individuals falling within specific harm categories (see Figure 20).

2. *Technology companies*

- Justifications for the banning of extremist individuals and organisations should be created and these should be consistent across platforms.
- The role of the GIFCT as an independent organisation will be instrumental in creating better inter-platform transparency on which organisations and individuals are problematic and why.²⁵³
- If a ban is to be implemented on individuals or organisations on the higher end of the extremism harm spectrum, technology companies will need to take care to ensure new accounts are not set up by the same individual or organisation, relying on technology to offset these attempts. In the offline space, extremist organisations often splinter into new names and forms, or advocate for less famous speakers to disseminate equally problematic ideas. Care must be taken that this is not echoed in the online space. Here, existing automated software to counter efforts of disinformation or fake ‘bot’ accounts can be employed.

3. *GIFCT*

- A new department should be created under the GIFCT research fund which focuses specifically on collating relevant information from publicly available court cases and convictions of extremist-related offenders using the internet to further their aims.
- As the report illustrates, several strategies used by those disseminating extremist content or functioning as extremist speakers on online platforms demonstrate similar characteristics (for example, the use of public broadcast and multiple platforms for greater dissemination, a history of offline violence or offences, multiple postings

²⁵² Social media companies already have some thresholds in place to measure harm; the report argues that these should be more consistent. For example, a three-strike rule on the part of the user, the dissemination of multiple pieces of harmful content, and user, police, and trusted flagger reports can all be used to link to online indicators.

²⁵³ ‘Next Steps for the Global Internet Forum to Counter Terrorism’, September 23 2019, available at: <https://newsroom.fb.com/news/2019/09/next-steps-for-gifct/>, last visited: 11 October 2019.

illustrating intent, a lack of remorse or limiting the significance of extremist statements); a historical record of those using platforms to spread hateful views would assist in flagging behaviour before it escalates into offline violence.

- In terms of a breakdown of responsibility, profiles graded in the 0–3.3 scale (low levels of harm) could be handled by internal departments in social media companies, in collaboration with existing information from the police.
- Profiles in the higher level of the harm spectrum (5.8 and above) could be handled by the new research department within the GIFCT, in the interests of sharing this information across technology companies for maximum effectiveness in prevention of offline harm.

4. Independent Regulator

- As the extremism grading scale evolves, and the behaviour of the individual in question becomes more harmful, faster action will need to be taken to remove content, users, or organisations from platforms.
- An independent regulator on online harms may assist with ensuring that a framework is implemented consistently across platforms, including lesser-known platforms where content and banned individuals and groups may migrate. The creation of a consistent indicator framework across platforms will assist in achieving this; the indicators in question should be dynamic and may evolve as more information becomes available from the offline space.

A multi-agency approach will be required to ensure checks and balances are maintained in the regulation of content and users. This will require participation from the Home Office, the DCMS (to map online harms caused by extremism and other types of harm), the police, and the CCE. It is hoped that the new and independent GIFCT organisation, as well as an independent regulator on online harms, leads in this with regular meetings to share case studies and approaches, and ensure that any database on extremist indicators and individuals is kept as up to date and relevant as possible, to prevent online harms from escalating into real-world violence. Anonymised case studies and examples of patterns of behaviour of extremism online could be published in an annual report, either by individual technology companies, the GIFCT department, or both.

Figure 20: Online Extremism Harm Thresholds

Harm Category	Analysis	Outcome - online	Outcome - offline
0-3.3	Offenders in this low category of online extremist harm tended to have strong networks of families or friends drawing them into committing extremism or terrorist-related offences. Individuals tended not to use public social media platforms. Offenders used encrypted apps.	Larger companies within GIFCT would benefit from working closely with police to ensure platforms are not used for terrorist financing or facilitation. Lesser-known platforms such as Surespot and Wickr could work with larger companies and police under the GIFCT.	All charges in this category were terror-related. Closer collaboration is needed to ensure any punitive measures such as orders not to post on social media are upheld.
3.4-5.7	Offenders in this category used both public and private platforms to communicate. There was an increased use of incitement to violence and the utilisation of social media sites as networking facilities to radicalise others and build relationships.	Platforms were used for different purposes: offenders using encrypted apps tended not to use public platforms to advocate extremist views. Those who fall in 3.4+ should be approached in a profile-driven and content-driven way. Extremist material that is shared on platforms for networking, relationship building, or audience amplification (paid sponsorship, advertising, or blue ticks) should be monitored when a threshold of indicators is crossed.	Glorification and/or justification of extremism occurred on private, rather than public, channels, feeding into a cycle of online radicalisation. Two offenders had a history of engagement with the authorities for extremism concerns. Technology companies could work closely with police to understand prior history of extremist behaviour.
5.8-8.1	There is a jump in public broadcast from the harm ranking of 5.8 and above. Also corresponding with the harm ranking of 5.8+ is an increase in the amount of material online which glorifies, and is associated with, a terrorist or extremist group.	Technology companies would benefit from having a transparent list of individuals and organisations not allowed on their platforms, to avoid the use of public features of their sites. Individuals in the harm category of 5.8+ focused on sharing and disseminating content; as a result, content removal will be required, as would limitations on the ability of individuals to sponsor and advertise content, reach new audiences, engage with existing audiences, and live stream.	Across the dataset, offenders tended to use hybrid options to communicate, including private and public mediums. Information sharing across platforms would allow for better monitoring of extremist speakers and content. An independent regulator could assist in benchmarking efforts to remove problematic content across platforms, and advise on various options to reduce accessibility of such content.

Harm Category	Analysis	Outcome - online	Outcome - offline
8.2-10.4	Offenders tended to have been previously warned about their behaviour, but continued to post content online.	An appeals process should allow individuals facing restrictions to present arguments repudiating their views, but care should be taken to ensure that the significance of past statements is not marginalised.	A similar appeals process has operated offline - see five qualitative case studies.
10.5-12.8	Offenders who fell in this category tended to have had fewer past warnings and indicated more remorse in courts.	Individuals would benefit from educative content and exposure to alternative content.	Escalation of harmful content will need to be reported in the offline space.
12.9-15.2	Offenders in this platform scored highly on all harmful indicators.	Individuals or organisations should be banned, with care taken to track new aliases and splinters. Consideration would need to be given as to how long bans are upheld for, and whether individuals can appeal after serving sentences offline.	Information on these individuals should be shared across platforms and with agencies offline where possible.

Title: "FREE TO BE EXTREME"

By Nikita Malik

© The Henry Jackson Society, 2020

The Henry Jackson Society
Millbank Tower, 21-24 Millbank
London SW1P 4QP, UK

www.henryjacksonsociety.org



This video has been removed for violating YouTube's policy on hate speech. Learn more about combating hate speech in your country.

[Learn more](#)



**CENTRE ON
RADICALISATION
& TERRORISM**