VOX
Pol

CHECK THE WEB

ASSESSING THE ETHICS AND POLITICS OF POLICING THE INTERNET FOR EXTREMIST MATERIAL

Ian Brown and Josh Cowls

CHECK THE WEB

ASSESSING THE ETHICS AND POLITICS OF POLICING THE INTERNET FOR EXTREMIST MATERIAL

Ian Brown and Josh Cowls
Oxford Internet Institute

About the authors

Professor Ian Brown is Professor of Information Security and Privacy at the Oxford Internet Institute. His research is focused on surveillance, privacy-enhancing technologies, and Internet regulation. He is an ACM Distinguished Scientist and BCS Chartered Fellow, and a member of the UK Information Commissioner's Technology Reference Panel.

Josh Cowls works on the Oxford Internet Institute's Accessing and Using Big Data to Advance Social Science Knowledge project, having previously completed an MSc at the OII.

Acknowledgements

The authors are extremely grateful to the speakers and participants of the associated VOX-Pol workshop, held in Brussels 19–20 January 2015, and to interviewees and all those that provided comments on a draft of this report.

ISBN: 978-1-873769-49-2

Designed and typeset by Soapbox, www.soapbox.co.uk

TABLE OF CONTENTS

INTRODUCTION

The last decade has seen growing societal concerns over the effect of violent extremist material hosted on websites, forums and social media. In Europe and elsewhere, police and security agencies are increasingly being asked to take action against it – in the form of monitoring such material, using it in investigations, and asking Internet companies to remove or block it.

A specific example of a police response is Europol's 'Check the Web' portal, which "enables competent authorities of EU Member States to share information on Islamist terrorist activities on the Internet via the secure Europol network and the Europol national units. Its aim is to create synergies between the Member States in the analysis of online Islamist terrorist activities".[1]

In January 2015, the EU's Justice and Home Affairs ministers suggested that "Internet referral capabilities, also through Check-the-web, could be developed within Europol to support efforts of Member States in detecting illegal content and improving exchange of information".[2] And on 1 July Europol officially launched their EU Internet Referral Unit "to combat terrorist and violent extremist propaganda".[3]

[1] Europol Review 2013, 3 September 2014, at www.europol.europa.eu/content/europol-review-2013

[2] EU Presidency, *Riga joint statement following the informal meeting of Justice and Home Affairs Ministers in Riga on 29 and 30 January*, at https://eu2015.lv/images/Kalendars/IeM/2015_01_29_jointstatement_JHA.pdf

[3] See www.europol.europa.eu/content/europol%E2%80%99s-internet-referral-unit-combat-terrorist-and-violent-extremist-propaganda]. The Justice Ministers' comments were in keeping too with a 2014 report from the UK Parliament's Intelligence and Security Committee that called for Internet companies to share much more user data with law enforcement agencies, and to take proactive steps to detect and report material relating to terrorism.

This report assesses the ethics and politics of policing online extremist material, using the normative framework of international human rights law, particularly the International Covenant on Civil and Political Rights, European Convention on Human Rights and the EU Charter of Fundamental Rights – whilst not conducting a legal analysis. It draws where appropriate upon interpretations by the UN Human Rights Committee, UN experts (such as the High Commissioner for Human Rights and special mandate holders), and regional human rights bodies and courts (such as the Council of Europe and the European Court of Human Rights).

The report looks at definitions of 'extremist material'; the types of monitoring and blocking being undertaken by government agencies and the private sector; and considers the roles of these key stakeholders, along with private individuals and civil society groups. It is based on a two-day workshop in January 2015 with thirty expert stakeholders from law enforcement and intelligence agencies, governments and parliaments, civil society, and universities. Short versions of ten papers were presented to stimulate discussion, following an open call for extended abstracts. These are available on the VOX-Pol website: http://voxpol.eu/.

The authors conducted seven follow-up semi-structured interviews with stakeholders from law enforcement, industry, government and civil society; and background policy analysis. The first author also co-organised a workshop on privacy and online policing with the UK's National Crime Agency in March 2015, and participated in three further workshops where the topics of this report were addressed: two on law enforcement use of communications data, and a third at the United Nations on the relationship between encryption and freedom of expression. Both authors are grateful for the assistance of interviewees, co-organisers, and workshop participants.

The report is produced by the EU-funded VOX-Pol Network of Excellence, and takes particular account of the network's development of semi-automated search for violent online extremist content and deployment of available tools for search and analytics, including text, video, sentiment, etc., currently employed in other domains for analysis of violent online extremist content. The network's focus is on making these tools freely available for research purposes to academics, but may also extend to others professionally tasked in this area (such as activists and law enforcement agencies). It is also centrally concerned with the ethical aspects of deployment of such tools and technologies.

NORMATIVE FRAMEWORK

WE USE AS a normative framework the broad principles of international human rights law, in particular the International Covenant on Civil and Political Rights (ICCPR), European Convention on Human Rights (ECHR) and EU Charter of Fundamental Rights (CFR). This framework has broad applicability; immense legitimacy given its ongoing development by the democracies since the Second World War; is familiar to and discussed at length by all European stakeholders; and is frequently referenced by international and European counter-terrorism policy-makers as essential to their efforts.

"Policies which are human rights compliant preserve the values the terrorists are trying to destroy, weaken support for radicalism among potential adherents, and strengthen public confidence in the rule of law."

Council of Europe's Commissioner for Human Rights

The United Nations Security Council resolved in 2014 that "respect for human rights, fundamental freedoms and the rule of law are complementary and mutually reinforcing with effective counter-terrorism measures, and are an essential part of a successful counter-terrorism effort",[4] while the EU's Justice and Home Affairs ministers stated in January 2015 that efforts to "detect and remove illegal content" must be "in full respect of fundamental rights".[5] Following the shocking shootings in Paris and Copenhagen in early 2015, the Council of Europe's Commissioner for Human Rights stated: "Policies which are human rights compliant preserve the values the terrorists are trying to destroy, weaken support for radicalism among potential adherents, and strengthen public confidence in the rule of law".[6]

The EU Member States are all parties to the ICCPR and the ECHR, while since the 2009 Lisbon Treaty, the CFR has been primary EU

[4] United Nations Security Council Resolution 2178 (2014), adopted on 24 September 2014.

[5] EU Presidency, note 2.

[6] Council of Europe Commissioner for Human Rights, 18 February 2015, at www.facebook.com/permalink.php?story_fbid=410537832455466&id=118705514972034

law; their counter-extremism efforts must therefore be in accordance with this framework. In total, 168 states are parties to the ICCPR, and 47 to the ECHR.

Other suggestions have been made for ethical frameworks to govern state analysis of online information for security purposes, notably that of former UK intelligence agency head and coordinator, Sir David Omand.[7] However, while these overlap significantly with the human rights law frameworks, they do not have the same democratic legitimacy or stakeholder familiarity.

Policing the Internet for extremist material can affect a number of the specific rights in the ICCPR, ECHR and CFR – particularly freedom of expression and privacy, but also freedom of association and assembly, and of thought, conscience and religion.

States have human rights obligations to protect the life and security of individuals under their jurisdiction,[8] justifying interference in other rights, so long as this interference meets the tests described below. Such restrictions "must be based on clear, precise, accessible and foreseeable legal rules, and must serve clearly legitimate aims; they must be 'necessary' and 'proportionate' to the relevant legitimate aim ... and there must be an 'effective [preferably judicial] remedy' against alleged violations of these requirements".[9] Restrictions must be subject to independent authorisation and oversight, and rights must be protected for everyone, without discrimination.

LAWFUL, NECESSARY AND PROPORTIONATE

The first general human rights law framework principle is that interferences with rights should be in pursuit of a legitimate purpose; necessary and proportionate to that purpose; and set out clearly in

[7] David Omand, *Securing the State*, London: Hurst & Co, 2010.

[8] Office of the United Nations High Commissioner for Human Rights, *Human Rights, Terrorism and Counter-terrorism*, Fact Sheet No. 32, July 2008.

[9] *The rule of law on the Internet and in the wider digital world*, Issue Paper published by the Council of Europe Commissioner for Human Rights, CommDH/IssuePaper(2014)1, 8 December 2014, p.10.

law, so that individuals and their political representatives can foresee and collectively determine the circumstances under which this may occur. As the UN Counter-Terrorism Implementation Task Force (CTITF) put it: “Whenever counter-terrorism measures… limit the full enjoyment of a human right, States must show that the measure was provided by law, and is both necessary and proportional”.[10]

Legitimate purposes are set out alongside each right in the ICCPR and ECHR – for example, in the case of the ECHR Article 8 right to privacy, “national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others” – and in article 52 of the CFR:

> Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others.

While elements of counter-extremism programmes that impact on human rights are often justified in the interests of national security, the CTITF has noted that such limitations “must be necessary to avert a real and imminent – not just hypothetical – danger to the existence of the nation, its territorial integrity or political independence”.[11]

Interferences must be necessary to achieve a legitimate aim, and proportionate to the aim – more significant reductions in harm might justify more significant interferences, but assessments must also take into account potential societal harms from interferences with

[10] United Nations Counter-Terrorism Implementation Task Force (CTITF) Working Group on protecting human rights while countering terrorism, *Basic Human Rights Reference Guide: Security Infrastructure*, 2nd edition, March 2014, §4.

[11] CTITF, note 10, §16.

rights.[12] Less intrusive alternative measures should be considered first. The international human rights law framework and its judicial interpretation provides guidance as to whether specific measures meet these tests, but some political judgment will still be required until a court gives a specific legal decision.[13]

It is well established by the European courts that legal powers to monitor and retain communications and other personal data without the knowledge of affected individuals interfere with privacy. In *Klass v Germany*, the European Court of Human Rights held that legislation allowing secret surveillance of communications "directly affects all users or potential users of the postal and telecommunication services… this menace of surveillance can be claimed in itself to restrict free communication through the postal and telecommunication services, thereby constituting for all users or potential users a direct interference with the right guaranteed by Article 8".[14]

More recently, the Court of Justice of the EU ruled that a legal obligation for public communications services "to retain, for a certain period, data relating to a person's private life and to his communications… constitutes in itself an interference with the rights guaranteed by Article 7 of the Charter", and that "the access of the competent national authorities to the data constitutes a further interference with that fundamental right".[15]

International sharing by states of personal data without a legal basis is problematic, and the European Parliament has called "on the Member States to refrain from accepting data from third states which have been collected unlawfully and from allowing surveillance activities on their territory by third states' governments or agencies which

[12] The ValueSec project developed decision support tools for security policy that enable cost-benefit analysis of security measures: www.valuesec.eu

[13] Fiona de Londras (ed.), *The Impact, Legitimacy and Effectiveness of EU Counter-Terrorism*, SECILE Consortium, November 2014, p.24 at http://secile.eu/wp-content/uploads/2014/11/SECILE_doc_amended.pdf

[14] *Klass and others v Federal Republic of Germany*, European Court of Human Rights (Series A, NO 28) (1979–80) 2 EHRR 214, 6 September 1978, §37.

[15] Court of Justice of the EU, Judgment of 8 April 2014 – joined cases C-293/12 and C-594/12, §§34–35.

are unlawful under national law or do not meet the legal safeguards enshrined in international or EU instruments".[16]

Alongside privacy, the Council of Europe has also stressed the importance of the Internet in supporting freedom of expression, and recommended that member states should not subject individuals to "any general blocking or filtering measures by public authorities, or restrictions that go further than those applied to other means of content delivery".[17]

Technical systems (including those used for Internet blocking and surveillance) often have high fixed costs but low marginal costs – leading to "function creep", where the use of technical measures to prevent significant harm enables much easier and cheaper use of the same measures to address less significant harms. Data from Internet monitoring can also be easily linked with other data sources – such as travel, financial, and medical records, and other forms of 'smart' surveillance such as digital CCTV with facial recognition tools.[18] Care is therefore needed by those assessing proportionality to avoid a creep towards "a publicly unacceptable level of overall surveillance"[19] and other interferences with rights, "the possession of a dangerous capability; and the overall damage to a medium that is of obvious intrinsic value beyond security".[20]

[16] European Parliament, Report on the US NSA surveillance programme, surveillance bodies in various Member States and their impact on EU citizens' fundamental rights and on transatlantic cooperation in Justice and Home Affairs, (2013/2188(INI)), 23 December 2013, §21.

[17] Council of Europe, Recommendation CM/Rec(2007)16 of the Committee of Ministers to member states on measures to promote the public service value of the Internet, 7 November 2007.

[18] Supporting fundamentAl rights, PrIvacy and Ethics in surveillaNce Technologies (SAPIENT), *Engaging stakeholders and civil society in the assessment of surveillance practices*, 26 September 2012, at www.sapientproject.eu/docs/D2.6%20-%20EngagingStakeholders%20%28submitted%2026%20Sept%202012%29.pdf, p.2.

[19] David Omand, Jamie Bartlett and Carl Miller, *#Intelligence*, London: Demos, 2012, p.41, at www.demos.co.uk/files/_Intelligence_-_web.pdf?1335197327

[20] *Ibid.*

Ensuring the active participation of citizens in determining counter-extremism policy means more than the public accessibility of legal codes and regular elections. As David Omand and colleagues have noted, “security and intelligence work in general is predicated not only on the public’s consent and understanding, but also on the active partnership and participation of people and communities. Serious and recognised damage to security occurs when the state’s efforts are not accepted or trusted”.[21]

This is a particular risk given the “flexibility… secrecy and urgency” of many counter-terrorism measures, which can “create greater distance, or ‘remove’ policies and processes from citizens’ influence. As a result, decision-making becomes less transparent and accountable and the link between the measures and democratic authorisation less tangible”.[22]

The SECILE project conducted extensive research with policy-makers and other stakeholders on what best creates democratic legitimacy for counter-terrorism legislation: they suggest “participation, deliberation, contestation, reviewability, and learnability in the development, implementation and review phases”.[23]

As well as widespread consultation over proposed legislation, they suggest that regular reviews are essential, not least since “changing socio-political conditions may result in adjustments in necessity and proportionality analyses”. Such reviews can also make use of experience in the implementation of measures, allowing a “more concrete and less speculative proportionality analysis [to] be undertaken than is possible at the *ex ante* stage”.[24]

[21] Omand, Bartlett and Miller, note 19, p.18.

[22] SECILE Consortium, note 13, p.13.

[23] SECILE Consortium, note 13, p.23. See also SAPIENT, note 18, and Privacy and emerging fields of science and technology: Towards a common framework for privacy and ethical assessment, *A Privacy and Ethical Impact Assessment Framework for Emerging Sciences and Technologies*, 25 March 2013, pp.87–110, at www.prescient-project.eu/prescient/inhalte/download/PRESCIENT_deliverable_4_final.pdf

[24] SECILE Consortium, note 13, p.29.

A long-running difficulty in democratic participation in security policy-making is that publicising some security measures can reduce their effectiveness. As a law enforcement officer told us: “The dilemma we have is if we reveal our methodology we inform behaviour and therefore lose our capability. It makes our job easier to apply covert methodology if the public doesn’t know about it. If made public everyone changes behaviour, especially the bad guys. There is a real balance to be struck on the amount of public information. We don’t tell people how we undertake conventional foot surveillance so people could spot it, as it is hardly covert at that point. This makes it a very difficult public debate to say we’d like to do all this stuff, if we explained people might think this reasonable, but its effectiveness is ruined”. Recognising this, the human rights framework allows a level of secret surveillance, but requires strong limits and safeguards for its use.

INDEPENDENT AUTHORISATION AND OVERSIGHT

A second general principle of the international human rights law framework is that state discretion in implementing measures that interfere with rights must not be ‘arbitrary’ or ‘unfettered’. Ideally this means that there should be independent judicial authorisation of such measures, “in a field where abuse is potentially so easy in individual cases and could have such harmful consequences for democratic society as a whole”.[25] Judicial authorisation in a state can also make international cooperation significantly easier when it wishes to request digital evidence from a second state that requires such control – for example, as the US does for communications intercepts.

Where individuals’ rights have been breached, the human rights framework requires that they have an effective remedy. This typically requires that remedies “be known and accessible to anyone with an arguable claim that their rights have been violated”, and that states undertake “prompt, thorough and impartial investigation of alleged violations”. Some states notify targets of surveillance once

[25] European Court of Human Rights, note 14, §56.

an investigation has concluded, allowing those that feel this was not justified to bring a claim.[26]

To be most effective, judicial control needs all of the elements of judicial independence. While there is a need for some operational details of investigations to remain confidential, secret hearings and judgments reduce public trust, as do departures from usual mechanisms for judicial appointments, a lack of opposing counsel in hearings, and non-availability of legal aid for less wealthy individual applicants. The European Parliament "[i]s adamant that secret laws, treaties and courts violate the rule of law".[27]

While not a disinterested party, a civil society interviewee who took part in a legal claim against the UK government in the UK's Investigatory Powers Tribunal (IPT) told us: "for advocacy to be successful, it requires a tribunal willing to listen! There were many times we had to go back and correct the tribunal's recollection of our argument, who at points seemed to be wilfully misunderstanding what it was our counsel was saying. A very frustrating experience, and not one we will forget when we make submissions in Europe on whether the IPT is A.6 compliant". And lawyer Matthew Ryder has noted the difficulties for claimants caused by the lack of financial assistance at the IPT, where there is "almost no legal aid and no compensation for legal costs".[28]

The UN High Commissioner for Human Rights suggests that effective oversight needs the involvement of the executive, judicial and legislative branches of government, as well as an independent civilian oversight agency.[29] The Council of Europe's Commissioner for Human Rights recommends "Member states ... ensure that effective democratic oversight over national security services is in place.

[26] Office of the United Nations High Commissioner for Human Rights, *The right to privacy in the digital age*, A/HRC/27/37, 30 June 2014, §§39–41.

[27] European Parliament, note 16, §11.

[28] Matthew Ryder, "Police sins of surveillance go far beyond the Lawrences and must be exposed", the *Guardian*, 9 March 2014, at www.theguardian.com/commentisfree/2014/mar/09/lawrences-sins-of-surveillance-exposed-inquiry-undercover-policing

[29] UN OHCHR, note 26, §37.

For effective democratic oversight, a culture of respect for human rights and the rule of law should be promoted, in particular among security service officers".[30] The European Parliament has suggested "the setting up of a high-level group to strengthen cooperation in the field of intelligence at EU level, combined with a proper oversight mechanism ensuring both democratic legitimacy and adequate technical capacity", and stressed that "the high-level group should cooperate closely with national parliaments in order to propose further steps to be taken for increased oversight collaboration in the EU".[31]

Effective oversight also requires courts and oversight committees that can understand the complex and fast-changing technological issues involved in Internet monitoring, and that are as open as possible to ensure public understanding. The European Parliament has resolved: "oversight of intelligence services' activities should be based on both democratic legitimacy (strong legal framework, *ex ante* authorisation and *ex post* verification) and adequate technical capability and expertise, the majority of current EU and US oversight bodies dramatically lack both, in particular the technical capabilities".[32]

EQUALITY AND NON-DISCRIMINATION

Human rights are just that; they protect everyone, not just citizens of a specific state. Most states (although notably not the US) accept that they must be accorded to anyone within states' "power or effective control" – including to individuals outside the territory of states but affected by their actions. This is critical to policing online extremist material, given the transnational nature of online activities.[33]

Article 26 of the ICCPR further states: "all persons are equal before the law and are entitled without any discrimination to the equal protection of the law". As the UN CTITF has noted,

[30] Council of Europe Commissioner for Human Rights, note 9, p.24.

[31] European Parliament, note 16, §62.

[32] European Parliament, note 16, §60.

[33] UN High Commissioner for Human Rights, note 29, §§32–36.

Human rights are just that; they protect everyone, not just citizens of a specific state. Most states (although notably not the US) accept that they must be accorded to anyone within states’ “power or effective control” – including to individuals outside the territory of states but affected by their actions. This is critical to policing online extremist material, given the transnational nature of online activities.

this is “crucial for effectively countering terrorism”,[34] given the damaging impact discrimination can have on communities and their relationships with governments.

Measures that have a disproportionate impact on specific communities or groups of individuals can “lead to further marginalisation, discrimination and, in extreme cases, radicalisation within affected communities”.[35] Measuring the societal impact of counter-terrorism programmes therefore needs an understanding of the impact on “both a range of societal groups and all relevant societal values”.[36]

The former UN Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, Martin Scheinin, has noted that measures that profile individuals and lead to increased police attention to those that match certain group characteristics can “contribut[e] to the social construction of all those who share these characteristics as inherently suspect”, leading to “a feeling of alienation among the targeted groups”. Scheinin warned that “the victimization and alienation of certain ethnic and religious groups may have significant negative implications for law-enforcement efforts, as it involves a deep mistrust of the police”.[37]

[34] UN CTITF, note 10, §5.

[35] UN CTITF, note 10, §10.

[36] SECILE Consortium, note 13, p.12.

[37] Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism (A/HRC/4/26), 26 March 2007, §§55–58.

THE ROLE OF THE PRIVATE SECTOR

International human rights law applies to states, not to private actors, although states do have some level of obligation to ensure that the actions of private parties do not "impinge on the human rights of individuals".[38] This obligation is one reason why the EU Member States have collectively developed their extensive data protection framework, which applies to public *and* private organisations – although not for national security purposes, which are outside the scope of EU actions. (The Council of Europe's Data Protection Convention[39] *does* apply to national security activities. While seven states have entered reservations to this requirement,[40] this will not be possible under the updated version of the Convention finalised in 2015).

Internet infrastructure and services are almost entirely operated by the private sector. From a user perspective, the terms and conditions of online services can have a greater immediate impact than national laws, since if breached then content can be removed and accounts deleted without any criminal justice or judicial procedures. States therefore often involve Internet companies in co-regulating online behaviour – enabling them to also take advantage of those companies' expertise in a fast-changing technical field, and to encourage companies' cooperation in implementing agreed rules.[41]

There is a risk, however, that informal cooperation – outside a statutory framework – transfers decision-making from states that are legally obliged to respect and protect human rights, to private organisations that are not. The Council of Europe Commissioner for Human Rights concluded in a December 2014 report that "rule of law

[38] Council of Europe Commissioner for Human Rights, note 9, p.63.

[39] Council of Europe, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108, 28 January 1981.

[40] Andorra, Ireland, Latvia, Malta, Romania, Russia and Macedonia. See List of declarations made with respect to treaty No. 108, at www.conventions.coe.int/Treaty/Commun/ListeDeclarations.asp?NT=108&CM=1&DF=07/09/2014&CL=ENG&VL=1

[41] Ian Brown and Christopher T. Marsden , *Regulating Code: Good Governance and Better Regulation in the Information Age*, Cambridge: MIT Press, 2013.

obligations ... may not be circumvented through ad hoc arrangements with private actors who control the Internet and the wider digital environment... Member states should stop relying on private companies that control the Internet and the wider digital environment to impose restrictions that are in violation of the state's human rights obligations".[42] The former UN Special Rapporteur on Freedom of Expression, Frank La Rue, similarly noted:

> given that [Internet] intermediaries may still be held financially or in some cases criminally liable if they do not remove content upon receipt of notification by users regarding unlawful content, they are inclined to err on the side of safety by over-censoring potentially illegal content. Lack of transparency in the intermediaries' decision-making process also often obscures discriminatory practices or political pressure affecting the companies' decisions. Furthermore, intermediaries, as private entities, are not best placed to make the determination of whether a particular content is illegal, which requires careful balancing of competing interests and consideration of defences.[43]

[42] Council of Europe Commissioner for Human Rights, note 9, pp.21–23.

[43] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, to the UN Human Rights Council, A/HRC/17/27, 16 May 2011, §42.

DEFINITIONS OF EXTREMIST MATERIAL

DEFINING ONLINE 'EXTREMIST' material that should be subject to police attention, while protecting freedom of expression and other rights, is a difficult and contentious task. The boundaries between 'extremist' material meriting police attention, and speech that in the words of the European Court of Human Rights may "offend, shock or disturb"[44] – but is still protected under freedom of expression rules – are extremely difficult to define in general terms.

Incitement to commit terrorist acts has been widely criminalised, as recommended by a number of international bodies, including the UN Security Council.[45] Hate speech is also commonly proscribed. Some states further criminalise 'glorifying', 'praising' or 'apologising for' terrorism.

Searching for evidence of criminal acts in a human rights-compliant manner is clearly a responsibility of police. But other state actions impacting on freedom of expression, privacy and other rights – such as monitoring or blocking access to legal but 'extremist' material – need to be carried out with great care to be compliant with the human rights principles outlined in the previous section.

PROVOCATION OF TERRORISM AND HATE SPEECH

There is a reasonable consensus amongst states and international organisations that inciting a terrorist offence, recruiting, and training for terrorism should be criminalised – while recognising that there is no comprehensive international agreement on the definition of terrorism, given the difficulties of protecting legitimate political actions and activities resisting state activities that are themselves contrary to international law.[46]

The Council of Europe's 2005 Convention on the Prevention of Terrorism requires parties to criminalise "public provocation to

[44] Handyside v The United Kingdom (1976) 1 EHRR 737 §49.

[45] United Nations Security Council, Threats to international peace and security (Security Council Summit 2005), S/RES/1264, 14 September 2005, §1.

[46] *Human Rights, Terrorism and Counter-terrorism*, note 8, pp.5–7.

commit a terrorist offence", "recruitment for terrorism" and "training for terrorism",[47] which was implemented by the EU in 2008.[48] The Convention requires this be done "respecting human rights obligations", "subject to the principle of proportionality … and should exclude any form of arbitrariness or discriminatory or racist treatment". The former UN Special Rapporteur on protecting human rights while countering terrorism, Martin Scheinin, recommended this provision as best practice.[49]

It is not always easy for prosecutors to prove intent with these offences. A law enforcement officer told us: "If you take the analogy of a grooming offence, inherently, there's a lot of difficulty proving intent of the grooming – someone just chatting and being friendly. What is the framework for intervening against someone grooming to go to Syria? What information is exchanged, and could it be interpreted other ways?"

Most European states have criminalised various forms of hate speech, and ICCPR §20(2) requires the proscription of "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence". The Council of Europe's Additional Protocol to the Budapest Convention on Cybercrime requires the criminalisation of the dissemination of "racist and xenophobic material" through computer systems, and encourages criminalisation of racist/xenophobic threats of a serious criminal offence, racist/xenophobic insults, and denial, approval or justification of genocide or crimes against humanity.[50]

[47] Council of Europe Convention on the Prevention of Terrorism, ETS No. 196, Warsaw, 16 May 2005, §12.

[48] Council Framework Decision 2008/919/JHA of 28 November 2008 amending Framework Decision 2002/475/JHA on combating terrorism, *OJ L 330*, 09/12/2008, p. 21–23.

[49] Australia: Study On Human Rights Compliance While Countering Terrorism, A/HRC/4/26/Add.3, 14 December 2006, §26.

[50] Council of Europe, Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, ETS No. 189, 28 January 2003, §§3–6.

In a recent review for the Council of Europe, Tarlach McGonagle suggested that criminal measures may be most appropriate for the "most egregious" hate speech, but other types may be better met with "educational, cultural, informational and other non-regulatory" measures.[51] Similarly, the 2012 Rabat Plan of Action, coordinated by the UN High Commissioner for Human Rights, recommends: "Criminal sanctions related to unlawful forms of expression should be seen as last resort measures to be only applied in strictly justifiable situations… legislation should be complemented by initiatives coming from various sectors of society geared towards a plurality of policies, practices and measures nurturing social consciousness, tolerance and understanding change and public discussion".[52]

GLORIFYING AND APOLOGISING FOR TERRORISM

Some states have gone further than criminalising public provocation to terrorism. Three European examples are France, Spain and the UK, which criminalise in specific circumstances glorification of terrorism, including "reckless" dissemination (UK), and "apologie du terrorisme" (France).

Spain has criminalised the "'praising or justification, through any means of public expression or broadcasting' of terrorist offences".[53] In the UK, §§1–4 of the Terrorism Act 2006 criminalises the

[51] T McGonagle, *The Council of Europe against online hate speech: Conundrums and challenges*, Background paper for Polish Government and Council of Europe conference "The hate factor in political speech – Where do responsibilities lie?" Warsaw, 18–19.9.2013, pp.4–6, 28–29.

[52] Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. Conclusions and recommendations emanating from the four regional expert workshops organised by OHCHR, in 2011, and adopted by experts in Rabat, Morocco on 5 October 2012.

[53] Article 578 of the Spanish Penal Code. Translation from Mathias Vermuelen, *Assessing Counter-Terrorism as a Matter of Human Rights: Perspectives from the European Court of Human Rights*, in Fiona De Londras & Josephine Doody (eds.) *The Impact, Legitimacy and Effectiveness of EU Counter-Terrorism*. London: Routledge, 2015.

distribution of a “terrorist publication” with the intended effect, or is reckless as to whether the conduct has the effect, of “direct or indirect encouragement or other inducement to the commission, preparation or instigation of acts of terrorism” or “provision of assistance in the commission or preparation of such acts”, including “matter which … glorifies the commission or preparation (whether in the past, in the future or generally) of such acts… [which someone] could reasonably be expected to infer … should be emulated by him in existing circumstances”.

The UN Human Rights Committee in its regular report on ICCPR compliance recommended the UK “consider amending that part of section 1 of the Terrorism Act 2006 dealing with ‘encouragement of terrorism’ so that its application does not lead to a disproportionate interference with freedom of expression”.[54] The UK government replied that these powers have very rarely been used, and that following an independent review and separate parliamentary consideration, it still considers the offence does not “directly or indirectly undermine freedom of speech”.[55]

Benjamin Ducol has described the evolution of similar provisions in French law. Shootings in Toulouse and Montauban in March 2012 were the first terrorist acts in France for over a decade. Then-president Nicolas Sarkozy proposed in response an offence of “habitually visiting websites that advocate terrorism or call for hatred and violence”. The National Digital Council asked to be consulted, and civil society groups such as Reporters sans frontières made general criticisms. While the Cabinet approved the bill, with a penalty of up to two years in jail and a fine of €30,000, the opposition Socialist Party refused to vote on the bill before the upcoming election.

Once elected, President Hollande’s government introduced a new anti-terrorism bill in October 2012 with a similar provision, and also extended Internet monitoring powers. A further bill in September 2014 included a range of Internet-related measures.

[54] UN Human Rights Council, *Concluding observations*, CCPR/C/GBR/CO/6, 30 July 2008, §26.

[55] United Kingdom, *State party’s report*, CCPR/C/GBR/7, 29 April 2013, §1031.

There was a broad campaign against these by civil society and the private sector, but the bill was adopted. Its main elements allow administrative filtering for websites advocating terrorism, without judicial oversight; *ex post* review by the *Commission Nationale de l'Informatique et des Libertés*; and punishment for the consultation of websites praising terrorism.

In November 2014 the "apologie du terrorisme" offence was moved from the press law to the criminal code as a "terrorism offence". This allows police to use special techniques that include online monitoring and surveillance; extends limitation periods, reduces procedural guarantees; and allows the use of preventive detention. The online offence is now subject to up to seven years' imprisonment and a fine of €100,000, while an offline offence can be punished with up to five years' imprisonment and a fine of €75,000. The government has also discussed further measures directly relating to social media.[56]

The focus of offences is moving from behaviours to opinions, with a transition from preventing violent action to preventing radicalisation. The Internet is treated as an exception, with a different regime to other media and a separation from hate speech laws.

Ducol suggests that the French example demonstrates trends that can also be seen in other countries. The focus of offences is moving from behaviours to opinions, with a transition from preventing violent action to preventing radicalisation. The Internet is treated as an exception, with a different regime to other media and a separation from hate speech laws. There are also elements of extra-judicial decision-making and elements of private policing by the Internet industry, along with questions of effectiveness.

[56] Benjamin Ducol, http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/ http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

Saltman and Russell have pointed out a difficulty with dealing with illegal extremist material: the importance of context. Since such material is often written, it requires interpretation by readers, which may not be consistent. Even imagery relating to or documenting terrorist activities can be used in different settings – for example, in media reports, and by counter-extremism practitioners. An example of the latter is the US State Department's online counter-speech campaign, #ThinkAgainTurnAway, which "uses IS content but changes the messaging so that it works against IS propaganda".[57] A law enforcement officer told us "judicial oversight is important to verify within the context of a message that something is an offence".

The Rabat Plan of Action suggests a six-part test for criminally prohibited incitement of hatred, taking account of context; the position or status of the speaker; intent (beyond "negligence and recklessness"); content or form; extent; and likelihood, including imminence.[58]
The UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression and the ACHPR Special Rapporteur on Freedom of Expression and Access to Information adopted a Joint Declaration in 2008 on defamation of religions, and anti-terrorism, and anti-extremism legislation, which states:

> The criminalisation of speech relating to terrorism should be restricted to instances of intentional incitement to terrorism, understood as a direct call to engage in terrorism which is directly responsible for increasing the likelihood of a terrorist act occurring, or to actual participation in terrorist acts (for example by directing them). Vague notions such as providing communications support to terrorism or extremism, the 'glorification' or 'promotion' of terrorism or extremism, and the mere repetition of statements by terrorists, which does not itself constitute incitement, should not be criminalised.

[57] Erin Marie Saltman and Jonathan Russell, *The Role of Prevent in Countering Online Extremism*, London: Quilliam, 2 December 2014, p.7.

[58] UN High Commissioner for Human Rights, note 52, §22.

The UN Human Rights Committee has similarly stated that restrictions on freedom of expression must establish “a direct and immediate connection between the expression and the threat”.[59] This suggests care is needed before other measures are taken against online threatening speech. There is otherwise a risk that legitimate political debate will be affected, “particularly within immigrant or minority communities”.[60]

NON-VIOLENT EXTREMIST MATERIAL

Beyond the categories of speech already described, which many states have proscribed by law, there is less consensus on what constitutes online ‘extremist’ material that should be policed – especially where it does not directly encourage violence. As a law enforcement officer told us: “It is difficult to define the extremist material that should be covered. Everyone has a sense, but when you ask anyone to define this they have difficulty. There is an internal debate to what extent we are acting as the thought police”.

A concrete example of online extremist material is the Internet postings of Anders Behring Breivik before he committed his atrocities in Norway. While much of it was hateful, it was not all by any means illegal hate speech.

Robindra Prabhu has described the national debate in Norway following this event, including an investigation by police and an independent commission, which concluded that this material could not have provided a warning of Breivik’s plans, even with more active monitoring by police.[61] An academic analysis by Jacob Asland Ravndal similarly concluded that Brievik’s online posts between 2002

[59] UN Human Rights Committee, *General Comment No. 34*, CCPR/C/GC/34, 12 September 2011, §35.

[60] International Commission of Jurists (ICJ). 2009. *Assessing Damage, Urging Action.* Report of the Eminent Jurists Panel on Terrorism, Counter-terrorism and Human Rights, p.130, at www.un.org/en/sc/ctc/specialmeetings/2011/docs/icj/icj-2009-ejp-report.pdf

[61] Robindra Prabhu, At http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

and 2011 were not unusual compared to other comments on mainstream news websites – although Breivik was able to gain important explosives knowledge from online sources, and may have solidified his views and become disconnected from his family through his online activities.[62]

A civil society interviewee told us: "Ministers and officials will always privately accept governments are a blunt tool to deal with the nuances of society, and yet they feel inherently they have a responsibility to protect, bound up in the social contract. For me legislation is not a good tool to try and manage what is a very complex phenomenon. That said, police say to ministers, until it's a criminal offence I don't have any tools to do this, other than to pop around and say to them this isn't terribly nice, and that's the challenge that the police find themselves in. This is where some groups and individuals are very good at staying just on the right side of a fine line. And so we're now in a bind because we accept that legislation is a blunt tool and may not be able to respond, but on the other hand the police don't have tools and capability to do that. That's when we ask parliamentary draftsmen to try and have an effect, lawyers can then interpret up to a point. But the state should be a reluctant intruder in people's lives. Often the principle is sound, the rationale for doing something is right. It is the way you go about it that causes all sorts of problems. If we believe that extremist material is wrong and we have defined that, we need to help police to get rid of it".

A civil society interviewee told us: "There are some categories of material that police are regularly monitoring – for example, harassment, causing alarm and distress, and between potential protestors – so I am OK with monitoring extremism in relation to specific acts of criminality, and concerns about public order and safety".

Akil Awan has described how the Internet has become the principal platform for dissemination of the culture of jihadism. Such material has changed significantly in last five years, adapting to a younger

[62] Jacob Aasland Ravndal, "Anders Behring Breivik's use of the Internet and social media", *Journal EXIT-Deutschland*, 2, 2013, at http://journals.sfu.ca/jed/index.php/jex/article/view/28

"selfie-generation" that is less likely to understand Arabic. It sacrifices coherence for cogency, celebrating deed over word and directed action over piety and learning, spreading a "pornography of violence". It is aimed at religiously illiterate individuals, perhaps with a criminal background, with a sense of purposelessness or boredom.[63]

Some states are aiming to challenge ideologies that provide a fertile ground for recruiting individuals to terrorism. The UK's counter-terrorism prevention strategy states: "People who accept and are motivated by an ideology which states that violence is an acceptable course of action are more likely to engage in terrorism-related activity... Challenging that ideology is therefore an essential part of a preventative counter-terrorism programme". A Dutch social programme gathers frontline professional reports of "'utterances that have not (yet) broken any law', for example radical ideological messages or the propagation of values deemed 'incompatible with those of Dutch society'".[64]

The UK's approach is to challenge such ideologies through debate, rebuttal, and prosecution where illegal statements are made – as well as preventing non-British citizen 'apologists for terrorism' from visiting the country, which has been applied to "extreme animal-rights activists and anti-abortionists, anti-Semites, Islamophobes and neo-Nazis, as well as people broadly associated with terrorist and other extremist groups".[65]

Challenging political ideologies is an area where the greatest care needs to be taken with measures that could interfere with freedom of expression, and the right to participate in public affairs (§25 of the ICCPR). The UN Human Rights Committee has held: "In order to ensure the full enjoyment of rights protected by article 25, the free

[63] Akil Awan, At http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

[64] Marieke de Goede and Stephanie Simon, "Governing Future Radicals in Europe", *Antipode* 45(2), pp.315–335, March 2013, at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8330.2012.01039.x/abstract

[65] UK government, *Prevent Strategy*, Cm 8092, June 2011, Chapter 8, at www.gov.uk/government/uploads/system/uploads/attachment_data/file/97976/prevent-strategy-review.pdf

communication of information and ideas about public and political issues between citizens, candidates and elected representatives is essential".[66]

> **"There has been a remarkable unwillingness of governments to clearly legislate for standards of extremist content, to define in clear legal language what is content of concern, and to have a clear legal process to assess content and to challenge the assessment. There needs to be due process and a right of appeal."**
>
> Industry interviewee

An industry interviewee told us: "There has been a remarkable unwillingness of governments to clearly legislate for standards of extremist content, to define in clear legal language what is content of concern, and to have a clear legal process to assess content and to challenge the assessment. There needs to be due process and a right of appeal. Some of these processes have developed so quickly that due process norms are not applied with Internet content – for example, YouTube has received takedown requests about material published by mainstream news orgs". They added: "Different governments have different definitions. It's better to have ideas out in the open and discuss them – this allows you to intervene before individuals become violently radicalised. The US is doing better job at this, and have been thinking about it longer. Governments need to have communications teams looking more carefully at how to use new platforms being used by extremists to try and prevent other people becoming radicalised. But this is a very hard social debate to have, since you have to admit that radicalisation happens in many places and needs a complex and nuanced response. Politics is bad at that".

A civil society interviewee told us: "There is a big definitional problem with extremism – it potentially outlaws huge categories of political speech e.g. anyone who wants to question the history of democracy, the rule of law, who might because they're so upset about the role of Britain and America in Iraq, in helping to create the status

[66] UN Human Rights Committee, *General Comment No. 25*, CCPR/C/21/Rev.1/Add.7, 27 August 1996, §25.

quo where ISIS has grown directly out of Iraq, the failure to handle Syria. Anyone who wanted to vent their anger would be labelled extremist under some definitions. This is hugely problematic for political expression, political of freedom, and freedom of thought".

The international human rights framework is clear on material that should not be restricted on the basis of national security concerns. The Johannesburg Principles, which draw together international and regional law and standards, have been regularly endorsed by the UN Special Rapporteur on Freedom of Opinion and Expression, and by the UN Commission on Human Rights and Human Rights Council. Principle 7 states this "includes, but is not limited to, expression that:

1. advocates non-violent change of government policy or the government itself;
2. constitutes criticism of, or insult to, the nation, the state or its symbols, the government, its agencies, or public officials, or a foreign nation, state or its symbols, government, agencies or public officials;
3. constitutes objection, or advocacy of objection, on grounds of religion, conscience or belief, to military conscription or service, a particular conflict, or the threat or use of force to settle international disputes;
4. is directed at communicating information about alleged violations of international human rights standards or international humanitarian law".

It also states that "No one may be punished for criticising or insulting the nation, the state or its symbols, the government, its agencies, or public officials, or a foreign nation, state or its symbols, government, agency or public official unless the criticism or insult was intended and likely to incite imminent violence".[67]

[67] *The Johannesburg Principles on National Security, Freedom of Expression and Access to Information*, Article 19, November 1996, at www.article19.org/data/files/pdfs/standards/joburgprinciples.pdf

ISSUES WITH 'RADICALISATION'

In the international human rights law framework, interferences with rights must be proportionate to the harm prevented, and utilise the least intrusive means for doing so. This requires a clear understanding of the nature of those harms, and the effectiveness of different measures in reducing them.

'Radicalisation' is a commonly used term in counter-terrorism discourse – amongst experts and in popular debates. To become a 'radical' – "A person who advocates thorough or complete political or social reform; a member of a political party or part of a party pursuing such aims"[68] – is of course entirely legitimate in a democracy. Policy-makers usually add (or imply) "to terrorism or violent extremism".[69] The Council of the European Union has tried to distinguish the former sense as 'radicalism'.[70] The participants at our workshop agreed that while radicalisation (in the latter sense) can be a problematic term, it is a useful and well-understood shorthand for those addressing the issue.

The nature of this type of radicalisation is heavily contested. How far can commonalities be found between the complex socio-economic situations, and history of Internet usage, of individuals that have gone on to commit violent, ideologically-inspired acts – and what does this mean for policing online behaviour?

The Internet plays an important part in reinforcing extremist ideology and facilitating terrorist activity,[71] but there are very few cases of individuals being autonomously radicalised. Most potential radicals

[68] *Oxford English Dictionary*, 2015.

[69] Radicalisation Awareness Network, *Preventing Radicalisation to Terrorism and Violent Extremism: Strengthening the EU's Response*, RAN Collection: Approaches, lessons learned and practices, 14 January 2014, at http://ec.europa.eu/dgs/home-affairs/what-we-do/networks/radicalisation_awareness_network/ran-best-practices/docs/collection_of_approaches_lessons_learned_and_practices_en.pdf

[70] Council of the European Union, *Media Communication Strategy: European Union Strategy for Combating Radicalisation and Recruitment through Effective Communication of EU Values and Policies*, 5469/3/07, 6 June 2007.

[71] UK government, note 65, §5.45.

How far can commonalities be found between the complex socioeconomic situations, and history of Internet usage, of individuals that have gone on to commit violent, ideologically-inspired acts – and what does this mean for policing online behaviour?

are young digital natives to whom 'new' media isn't at all 'new'.[72] Real-world connections and experiences and peer groups seem to be most important in introducing individuals to extremist ideologies,[73] although the Internet can act as an 'echo chamber' to confirm existing beliefs.[74]

According to Akil Awan, taking UK media discourse as an example, 'radicalisation' is a relatively recent concept, and it is worth considering how it has come about. Media analysis shows that it was barely used before the July 2005 bombings in London, but then became a permanent part of the media landscape. There are spikes in usage with attacks, but also alongside discussions of security and legislation – and even academic conferences.[75] An analysis by Sedgwick found that usage of the term had tripled in English language media between 2003 and 2007.[76]

The concept seems to have been created to explain something inexplicable – why four young relatively integrated British men would blow themselves up – to an increasingly anxious and accusing public. Awan argues the term is obfuscating and packages all uncertainty into one word (see Figure 1), and that there seems little in common between individuals deemed 'vulnerable to radicalisation' other than being young, male and Muslim.

[72] Awan, note 63.

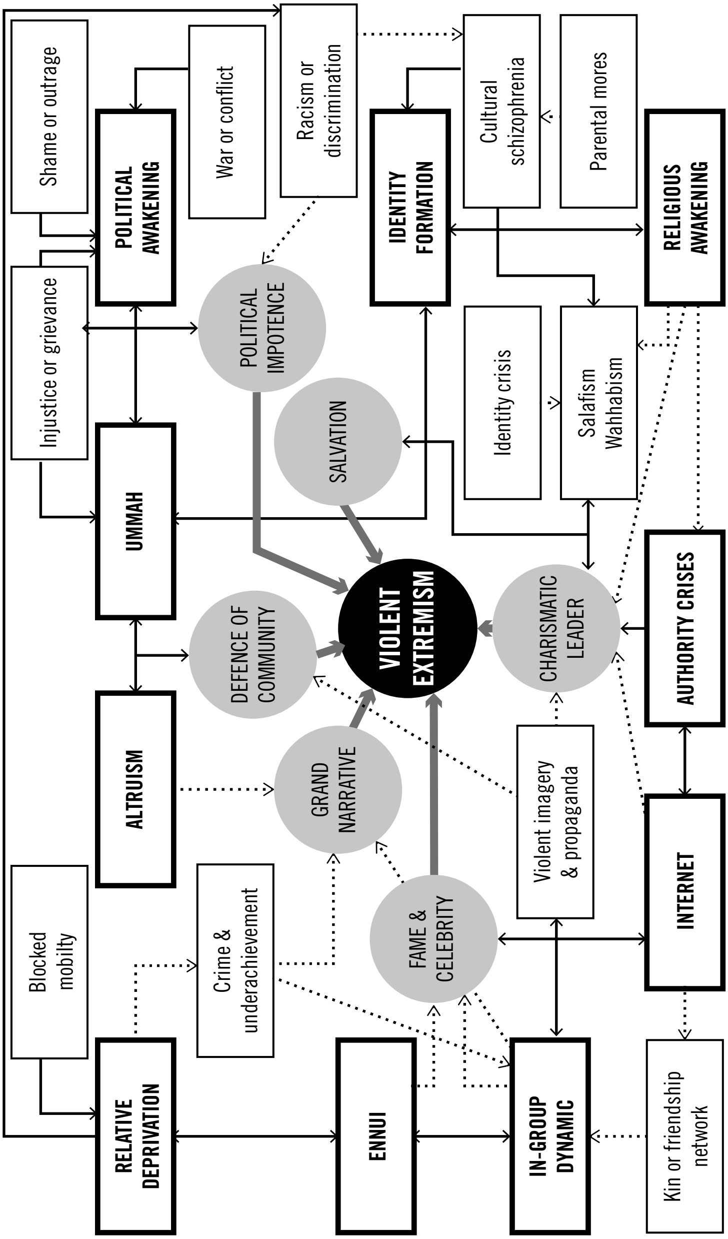
[73] Saltman and Russell, note 57, p.2.

[74] Ines von Behr, Anaïs Reding, Charlie Edwards and Luke Gribbon, *Radicalisation in the digital era: The use of the internet in 15 cases of terrorism and extremism*, RAND Corporation, 2013, at www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR453/RAND_RR453.pdf

[75] Awan, note 63.

[76] M Sedgwick, "The concept of radicalisation as a source of confusion", *Terrorism and Political Violence* 22(4), 2010, pp.479–494.

Figure 1: Pathways to violent extremism (Awan 2013)



The concept also focuses on forces that constrain people in various ways, de-emphasising their individual agency. De Goede and Simon characterise the "threat of incipient violence" as "the means by which early interventions in this broad field of potential radicals are justified and depoliticised as relationships of care".[77]

A 2008 analysis by the UK's Security Service concluded there is no "typical pathway to violent extremism". A typical "British terrorist" is "demographically unremarkable", legally in the country, ethnically diverse, a religious novice, and often married with children. The level of mental illness or pathological personality traits is no higher than in the general population.[78] A civil society interviewee added: "There is no rhyme or reason why some individuals will do what they are going to do, so every route is going to be slightly different".

Figure 1 on page 36 shows Awan's model of the complex factors and pathways that have been identified relating to violent extremism.

Beyond this, there is little understanding of the relationship between thoughts, words, actions, and the holding of non-violent extreme views. Religion is often a motif rather than a motivation of violent extremists. Yusuf Salwar and Mohammed Ahmed, who pleaded guilty to terrorism offences in the UK in July 2014, had purchased *The Koran for Dummies* and *Islam for Dummies* before travelling to Syria.[79] 'Extreme' online behaviour won't necessarily translate to the real world. More clarity is needed if online policing interventions are to be effective.[80]

There is also a broader impact on political debate and freedom of expression through the use of terms such extremism, radicalisation and terrorism, which "shape the limits of discourse through their implicit narratives and assumptions". Security threats are socially

[77] De Goede and Simon, note 64.

[78] Alan Travis, "MI5 report challenges views on terrorism in Britain", the *Guardian*, 20 August 2008, at www.theguardian.com/uk/2008/aug/20/uksecurity.terrorism1

[79] Vikram Dodd and agencies, "Two British men admit to linking up with extremist group in Syria", the *Guardian*, 8 July 2014, at www.theguardian.com/world/2014/jul/08/two-british-men-admit-linking-extremist-group-syria

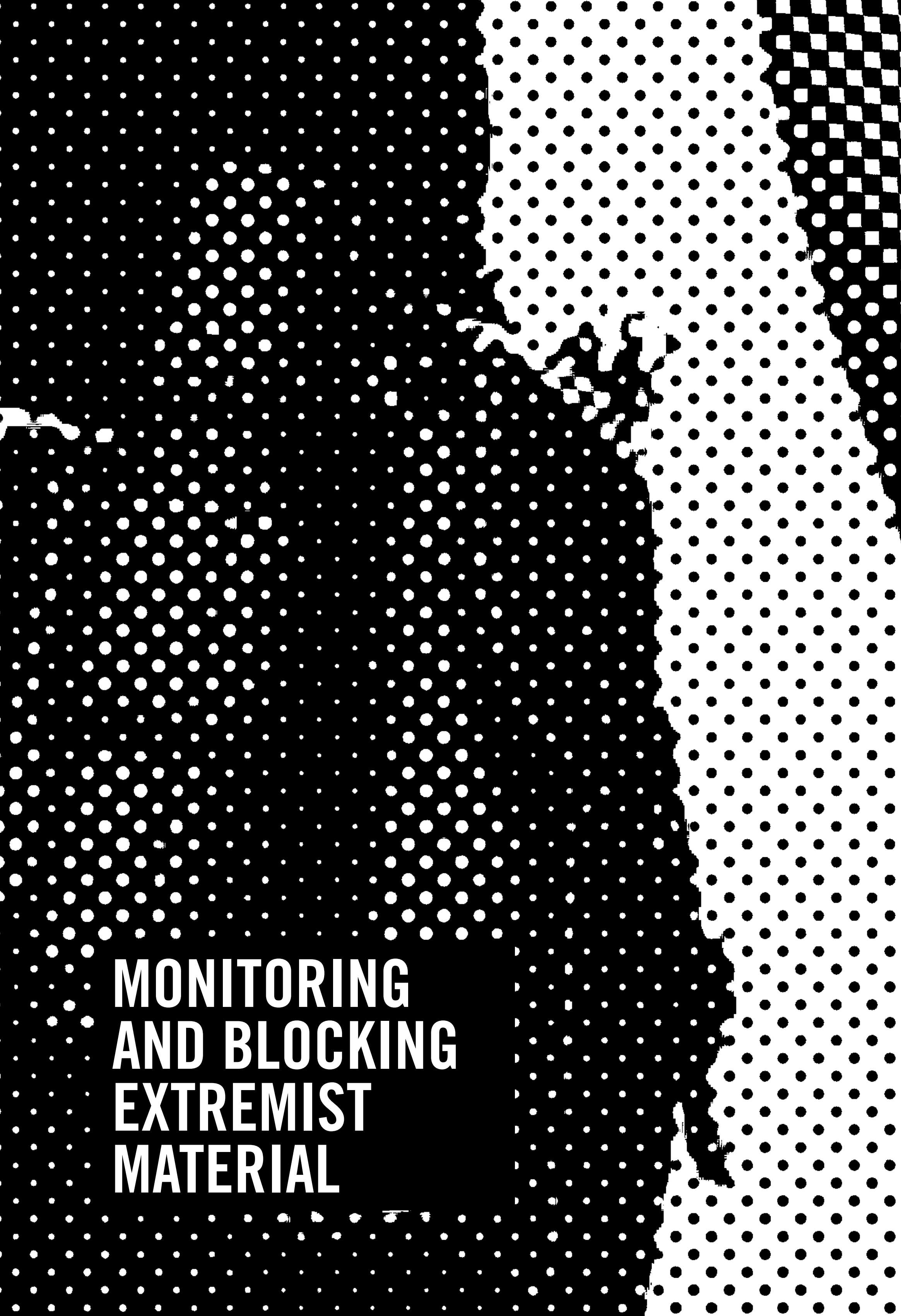
[80] *Ibid.*

and politically constructed, and not always linked to clear evidence of different types and levels of violence that exist in societies. Monitoring of 'extremist' ideologies can create a "culture of self-censorship" in discussion of radical political change – "to talk about foreign policy, for example, can now attract the suspicion that you might be labelled an extremist if you are in a school, in a youth club or in a university student society".[81]

As De Goede and Simon note, "The problematisation of radicalisation recasts political motivations and societal problems – including disenfranchisement and foreign policy disagreement – from problems deserving of political attention to *signs of potential danger* on the radicalisation trajectory, while disentangling itself from a serious engagement with potentially explosive political agendas" [emphasis in orginal].[82]

[81] Arun Kundnani, *Counter-Terrorism Policy and Re-Analysing Extremism*, Institute of Race Relations, 12 February 2015, at www.irr.org.uk/news/counter-terrorism-policy-and-re-analysing-extremism/

[82] De Goede and Simon, note 64.



MONITORING AND BLOCKING EXTREMIST MATERIAL

THERE IS A broad recognition in law enforcement and security agencies of the possibilities of using social media analysis – including for "understanding social resentments, grievances and radicalisation, and the identification of specific criminal intent or individuals... more rapid identification of events than traditional reporting mechanisms... [and] facilitating a faster, more effective, and more agile emergency response".[83]

These possibilities are particularly relevant to groups such as ISIS, which is aspirational, rather than groups such as Al Qaeda, which is conspiratorial. The former's online communications are full of calls to action for individuals, while the latter is much more closed, vetting individuals before they are allowed closer to the group.[84]

As Ducol notes, there is extremist content on websites, which also appears in online forums, chatrooms, via private communications tools such as WhatsApp, and many other venues.[85] As an example, Berger and Morgan estimated that in one eight-week period in late 2014, there were at least 46,000 Twitter accounts supporting ISIS.[86]

Policy-makers and law enforcement officials recognise the impact on human rights of "the general rise of information systems that have vast capacities to capture, stockpile, retrieve, analyse, distribute, visualise and disseminate information".[87] There is limited public understanding of the possibilities for surveillance this creates, both by government and companies; the potential for function creep; and the impact of possible data breaches. Even targeted surveillance

[83] Omand, Bartlett and Miller, note 19, p.24.

[84] The SecDev Group, *Spectral Sentinel: Advanced analytics for situational awareness and early warning of violent extremism*, 2013–2014.

[85] Ducol, note 56.

[86] J.M. Berger and Jonathon Morgan, *The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter*, The Brookings Project on U.S. Relations with the Islamic World, Analysis Paper No. 20, March 2015, p.7, at www.brookings.edu/~/media/research/files/papers/2015/03/isis-twitter-census-berger-morgan/brookings-analysis-paper_jm-berger_final_web.pdf

[87] Omand, Bartlett and Miller, note 19, p.32.

of individuals suspected of illegal behaviour will cause collateral intrusion into the lives of their contacts.[88]

McCarthy notes that appropriate and legitimate use of these tools depends on law enforcement agency experience with processing data, and having the resources to assess the merits of results. A "technology-push" approach leads to "short cut policing" without regard to the social impact or creation of distrust and insecurity – power relationships are disregarded to the detriment of greater societal security.[89] A risk is that "humans are increasingly relegated to the role of second-level decision-makers, with a range of potential discomforts and negative impacts for individuals subject to these systems… There is no doubt some surveillance yields social benefits, but equally there is no doubt that those controlling surveillance systems gain more power over those surveilled and targeted".[90]

Profiling and surveillance linked to specific communities can lead to resentment, damaging the good community relations with police that are vital to ongoing counter-terrorism cooperation. This has been a problem in some UK Muslim communities, with claims of discriminatory monitoring. While a government review found "no evidence to support these claims", it recognised that counter-extremism programmes "must not be used as a means for covert spying on people or communities. Trust… must be improved".[91] Covert activities to influence individuals or groups have a particular risk of damaging this trust.

One mechanism used by several European states to draw a clear border between counter-extremism and counter-terrorism programmes is having the former managed by local rather than national authorities. Local municipalities are also thought to be closer to vulnerable individuals' day-to-day lives, making detection and response easier.[92]

[88] *Ibid.*

[89] Sadhbh McCarthy, at http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

[90] SAPIENT consortium, note 23, p.3.

[91] UK government, note 65, p.6.

[92] De Goede and Simon, note 64.

One example are Information Houses in Dutch cities, from which front-line professionals such as teachers and social workers can get advice on concerns they have about individuals "at risk or in the process of radicalisation". The municipality remains responsible unless individuals are undertaking acts preparatory to terrorist actions, such as "recording a video testament, purchasing potentially explosive substances, [or] obtaining blueprints", when the police will take over.[93]

Another example is the UK's multi-agency Channel partnership, where a wide range of bodies identify individuals "at risk", including "local authorities; police [organised in the UK on a local basis]; youth offending services; social workers; housing and voluntary groups".[94] One of a number of indicators for Channel referral is "attempts to access or contribute to violent extremist websites".[95]

De Goede and Simon noted that counter-radicalisation programmes frequently use a language of "care" rather than policing, due to the "intimate and situated" interventions undertaken. Front-line professionals seem more comfortable to "signal concerns" than "spot radicals". An example is the Dutch Nuansa programme, which focuses on countering radicalisation with "Hand & Heart" as well as "Eyes & Ears".[96] But as Prabhu has also noted, while the line at which individuals move from being "at risk" to "a risk" can be difficult to draw clearly, it can have very significant implications for the individual concerned.[97]

[93] Yousiff Meah and Colin Mellis, *Recognising and Responding to Radicalisation, Considerations for policy and practice through the eyes of street level workers*, Amsterdam: The RecoRa Institute, 2008, p. 31, at www.recora.eu/media/the%20recora%20report.pdf

[94] UK government, note 65, §9.10.

[95] *Ibid.*, §9.11.

[96] De Goede and Simon, note 64.

[97] Prabhu, note 61.

PUBLIC MEDIA

A very large amount of content is made publicly available online, and is potentially accessible to law enforcement monitoring extremist material. As well as social media, particularly those such as Twitter and YouTube that are used largely for one-to-many communication, this includes public websites, blogs, and archives of large discussion groups and forums.

A civil society interviewee told us publicly available media "is where we are able to identify how extremists are using social media sites to describe their group, history, organisation, why they believe in violence, why they are anti-state, whatever it may be. That is now in the public domain, accessible to everyone with an Internet connection, and being indexed by search engines. There's then a question what impact that has on society, and for a large percentage of people, they won't see it, it won't have an impact. It will only have an impact when the mainstream media pick it up. However, those who are curious, those who come across it by accident or by design, are potentially individuals who are vulnerable. We have plenty of examples from police and others of school kids looking at stuff online, all in the public domain, which may have an impact on them. There is a spectrum of impact – some will have none, some is just an amusing highlight, some is forwarding a message that shows the beheading of an individual in Syria, and nothing may come of it".

Without targeting of individuals, monitoring such material can provide a useful level of 'situational awareness' for security agencies and the general public. For example, the Norwegian Board of Technology determined following the Breivik attacks that a simple semantic analysis of Twitter data can in principle provide a valuable complement to the situational awareness of emergency response agencies in large scale crisis situations. They also found that social media were important information dissemination tools for the general public during the attacks, notably for the people on Utøya Island. At least 37 Utøya survivors have reported that they were active on social media during the shootings.[98]

[98] Prabhu, note 61.

Omand et al. suggest that social media analysis can also help police and policy-makers understand the shifting positions and key concerns on issues within violent groups and networks, identifying "hot topics" and how groups react to specific events.[99] Berger and Morgan found that they were able to follow the propagation of new media material, themes, and issues across a network of ISIS supporters on Twitter.[100]

Besides Europol's Check The Web programme, there is no coordinated activity at the EU level on open source monitoring, but an EU official told us that "every single national intelligence agency within Europe will engage in this, at an increasing pace, especially in the wake of the Charlie Hebdo attacks. The level of technical savviness and use of analytics very much depends on the Member State. This could range from a group of 10 people checking known Twitter accounts and trying to follow known suspects on Twitter and Facebook, on who previous intelligence exists – such as border agency records of individuals going to Syria or Iraq. Some countries allow officers to 'friend' a suspect on a social media platform to take a close look at their profile. There is also a difference whether they can friend somebody on FB to look at closed profiles. The large Member States do much more sophisticated analysis, with greater resources and more technical savvy".

"People can see social media analysis as a panacea, but from a psychological standpoint, presented identity is no more than a snapshot of what users want to portray to other people – so how accurate will any aggregate analysis be? It needs interpretation. We have to be very careful about any legislative framework built around this problem."

Law enforcement officer

However, a law enforcement officer told us: "People can see social media analysis as a panacea, but from a psychological standpoint,

[99] Omand, Bartlett and Miller, note 19, p.25.

[100] Berger and Morgan, note 86, p.55.

presented identity is no more than a snapshot of what users want to portray to other people – so how accurate will any aggregate analysis be? It needs interpretation. We have to be very careful about any legislative framework built around this problem". Berger and Morgan's analysis found that "The most active and visible accounts contain more noise, and their content is more carefully stage-managed by ISIS and its adherents".[101]

The fact that content is available on public platforms does not mean there is a social consensus all of it should be considered as public. Much of what individuals say online is neither strictly private, nor strictly public – while individuals are sharing content with the world, they are directing it at a small subset of the global population. Media can contain 'metadata' such as location and timestamps that may not be obvious to those posting it online, but which enables powerful monitoring capabilities such as tracking an individual over time.[102] Social media and other websites frequently change their privacy policies, meaning users may not realise what is and is not available beyond small friendship circles.

Nor is it the case in the European data protection and privacy legal framework that publicly available personal information can be used for any purpose, or individuals monitored in public spaces with no safeguards. The European Court of Human Rights held in *Rotaru v Romania* that "public information can fall within the scope of private life where it is systematically collected and stored in files held by the authorities. That is all the truer where such information concerns a person's distant past".[103]

An EU legal expert told us: "a 'public figure' is someone who willingly enters the public arena. Twitter is such an area, since in general users want their opinions to be heard. But even if you willingly enter a public arena you have a certain degree of privacy. To what extent differs on [a] case-by-case basis. If you are not a suspect in a terrorist

[101] Berger and Morgan, note 86, p.55.

[102] Prabhu, note 61.

[103] *Rotaru v Romania*, European Court of Human Rights Application no. 28341/95, 4 May 2000, §43.

investigation, but there is a police officer with a hunch you are involved in terrorist activities, and that person goes every day multiple times to your Twitter feed and checks your posts, that is a bridge too far. You don't expect a police officer to follow you everywhere you go in a public road checking everything you say. The time criterion is important – is monitoring happening for [a] long period? This is ethically much more problematic than going one day to [a] Twitter profile and checking the last two weeks of tweets".

After the Breivik murders, the Norwegian Board of Technology asked Norwegians if police should be monitoring open social media platforms. While 40% of respondents thought so, 40% said it would stop them using the types of words they imagined might be monitored online.[104]

A civil society interviewee told us: "It is unacceptable to monitor political speech, there has to be [a] much higher threshold for 'extremism'. Anyone that wants to question the status quo is potentially an extremist – that should mean they are monitored".

Police need to respect these nuances if they are to build a trusting relationship with the public in their use of online material, as well as comply with human rights principles. The Rabat Plan of Action notes: "The importance of the media and other means of public communication in enabling free expression and the realisation of equality is fundamental... New technologies – including digital broadcasting, mobile telephony, the Internet and social networks – vastly enhance the dissemination of information and open up new forms of communication".[105]

Omand et al. noted the risk "that the unregulated large-scale collection and analysis of social media data will undermine confidence in, and therefore the value of, this space". They suggested it should not be used to identify individuals, only to gather aggregated and anonymised data; should not be used as a means of criminal investigation; and should only access freely available information – "such

[104] Prabhu, note 61.

[105] UN Office of the High Commissioner for Human Rights, note 52, §28.

as noting an increase in social media communications in a specific area where demonstrations are taking place".[106]

Before going further than this, they suggest the level of privacy interference means that police should seek independent authorisation; for example, in the UK legal framework, that "the close following of an individual's Facebook page during the course of an investigation has similarities to [visual] surveillance as 'authorisable' by a senior police officer".[107] The Norwegian Board of Technology developed similar recommendations – that police should use only open information (and be clear about its definition); initially only use aggregate data; only identify individuals in specific investigations; and build in privacy preserving tools like face blurring.[108]

A law enforcement officer told us they broadly agreed with this analysis, although they would pay more attention to "the intent of person posting material, so if you're looking at a specific individual who is trying to gain maximum exposure of information, that is different to following someone down a street. Even if police are actively monitoring an individual account that doesn't necessarily need that authority. [UK surveillance law] is about the person and engaging with that person in a covert matter, which is specific to circumstances – about necessity and proportionately of that activity, on that day, at that time, for that purpose. So I'm very nervous about saying 'monitoring Twitter needs x permission.' That said, authorisation is just an independent, objective, confirmation of someone's assessment – whether at senior police officer or Secretary of State level, or anything in between – saying 'I've read that business case, I agree or disagree'".

The officer also distinguished covert and overt monitoring online, for example even when "visiting a Twitter account from a [police] computer, the user won't know about it. That is not like walking out on the street in a police uniform".

[106] Omand, Bartlett and Miller, note 19, p.31.

[107] *Ibid.*

[108] Prabhu, note 61.

Another issue for policing raised in our interviews was the difficulty of distinguishing locally-based individuals posting genuinely threatening material, from overseas posters who may have little connection with the local jurisdiction, and other individuals posing little threat. In Berger and Morgan's demographics dataset, only 1.5% of users had enabled location in one of their 200 most recent tweets, although some put location information of variable quality in their profiles.[109]

A law enforcement officer told us: "Of one thousand social media followers of a terrorist group, how many are journalists, undercover officers, foreign governments, actual or wannabe extremists, or sad 16 year olds? Is it the best use of our time to go around chasing a load of Twitter handles? Yes, there is legislation – can we enforce that on an extra-jurisdictional basis? Probably not, so we would be wasting our time. We are thinking hard about how you identify people to make an assessment of whether we should be worried". Berger and Morgan undertook this type of semi-automated analysis of ISIS supporters using social network analysis and (time-consuming) individual account review.[110]

While officers were enthusiastic about the potential of tools to reduce the burden of online policing, they felt that automated analysis tools currently have "Lots of potential, but zero capability. Technology needs to catch up with the thinking to be effective. Commercial sentiment analysis of broadcast social media at the moment is too rudimentary and is only one piece of a jigsaw. We wouldn't make decisions based on sole sensor activity. Individuals need to be saying the right or wrong things publicly – that sometimes happens, but not always, it tends to be a public perception based on publicity after an event. The tools need to be quite extensively tested to say whether this is worthwhile doing, or is churning up hundreds of dead intelligence leads that aren't worth looking at. If research starts to look at behavioural science and with confidence can say

[109] Berger and Morgan, note 86.

[110] *Ibid.*

somebody chooses to say certain things at certain times for certain purposes, that shows something in particular, of course we would be interested. Effectiveness is critical – we have finite resources and lots of work to do. Be careful with what you wish for".

The effectiveness of such tools is important for their proportionality. An EU legal expert told us that automated analysis tools can sometimes produce "so many false positives and negatives that they become almost impossible to use. It's a typical thing that adds more hay to the haystack instead of making it easier to find a needle in haystack. This discussion has been going on since before 9/11, when plotters were using code words for terms such as 'bomb'. How you spot these are important. It's not about a word or the content of a word, but the frequency of a specific word showing up in a conversation within a known social network of suspected terrorists. Then text analysis could be interesting, perhaps indicating a bomb or a planned attack. Sentiment analysis is mainly useful if you are on known jihadi forums. You already know these people are not there to discuss the weather, they are there to discuss ideology, and so you can use sentiment analysis to check the evolution of a poster on a forum – whether they are becoming more aggressive in tone or more violent. But there's a big step from moving from radicalising online to committing a violent act. It's very hard to detect this with sentiment/text analysis tools. There is a distinction in proportionality between detecting groups and content. It depends at which step of the investigative process you are deploying these tools. If you already have one clear terrorist suspect on the basis of human or other intelligence sources, you try to map out a person's network, do sentiment analysis, on the network of this particular person then it's a completely different story to trying to identify e.g. Pakistani immigrants living in London aged 18–30, that is much more ethically problematic".

TJ McIntyre has identified some of the human rights issues raised by takedowns and filtering of online material outside a public law framework, particularly for freedom of expression and privacy, including a lack of transparency, accountability, and effective remedies when mistakes are made.[111] The lack of a foreseeable legal basis for filtering is incompatible with Article 10 of the European Convention on Human Rights[112] – it is not "voluntary" to end-users, whether or not on the "public estate". The law must indicate with sufficient clarity the scope of any discretion to block and the manner of its exercise. As a European Commission staff working document noted:

> the adoption of blocking measures necessarily implies a restriction of human rights, in particular the freedom of expression and therefore, it can only be imposed by law, subject to the principle of proportionality, with respect to the legitimate aims pursued and to their necessity in a democratic society, excluding any form of arbitrariness or discriminatory or racist treatment.[113]

Yaman Akdeniz noted that blocking was considered in initial proposals for the Check the Web project, with "numerous Internet sites in a wide variety of languages… monitored, evaluated and, if necessary, blocked or closed down". However, later proposals said, "Member States will not be obliged to monitor, interrupt or shut down specific Internet sites".[114]

Some violent extremist groups have become very effective at negating the impact of takedowns and blocks, with affected content

[111] TJ McIntyre, At http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

[112] *Ahmet Yildirim v Turkey*, European Court of Human Rights Application no. 3111/10, 18 March 2013, §59.

[113] European Commission Staff Working Document, Accompanying document to the proposal for a Council Framework Decision amending Framework Decision 2002/475/JHA on combating terrorism: Impact Assessment, 14960/07 ADD1, Brussels, 13 November, 2007, p.29.

[114] Yaman Akdeniz, *Study of legal provisions and practices related to freedom of expression, the free flow of information and media pluralism on the Internet in OSCE participating States*, OSCE Office of the Representative on Freedom of the Media, 15 December 2011, p.143, at www.osce.org/fom/80723%20

"reappearing on a variety of platforms repeatedly".[115] One analysis even suggests that avoiding such 'negative' measures has become a process that energises supporters of such groups, with content moving like "swarms" across and between different platforms.[116]

The Council of Europe recommends that if filtering is to be applied, it should only be to specific content identified by a national authority, with an impartial body to review decisions.[117] The Council of Europe's Commissioner for Human Rights has stated that "proposals to make internet service providers responsible for taking down content that incites to terrorism without any judicial review are highly problematic".[118] And a European Commission staff working document noted the speedy reappearance of closed-down websites, circumvention of blocks, and difficulty of blocking content through other services such as peer-to-peer networks.[119]

McIntyre also noted the ease with which blocking technologies can be repurposed as tools of surveillance, intelligence gathering and prosecution. This has already been seen in the US with tools used to block known child abuse images, some of which now provide automated reports to law enforcement.[120]

SEMI-PRIVATE SPACES AND PRIVATE COMMUNICATIONS

While debates about violent online extremist material tend to focus on deliberately public material, much publicly accessible online material is aimed at a smaller audience, and contains information posters

[115] Saltman and Russell, note 57, p.9.

[116] Jamie Bartlett and Ali Fisher, 'How to beat the media mujahideen', *Demos Quarterly*, 5 February 2015, at http://quarterly.demos.co.uk/article/issue-5/how-to-beat-the-media-mujahideen/

[117] Council of Europe, *Declaration on freedom of communication on the Internet*, Adopted on 28 May 2003 at the 840th meeting of the Ministers' Deputies.

[118] Council of Europe Commissioner for Human Rights, note 6.

[119] European Commission Staff Working Document, note 113, pp.29–30/41.

[120] McIntyre, note 111.

might consider has a degree of sensitivity.[121] Twitter is full of conversations between individual users, and it is easy to find online archives of small discussion groups and newsgroups, join public Facebook groups, and observe information being provided over peer-to-peer file-sharing systems.

Some of this information can be useful for policing extremism. Omand et al. note that “Groups like the [far-right extremist] English Defence League use sites like Facebook to plan and organise their demonstrations, and access to such data could be a vital source of information”.[122] It is also amenable to the same automated analysis as intentionally public online material, although it is even more likely to present challenges of community-specific language and norms. However, since the level of interference with users’ privacy, freedom of expression and association is greater, the benefits and safeguards of such analysis need to be correspondingly higher to be justified.

Beyond this in terms of intrusiveness is access to intentionally private materials. This includes ignoring restrictions on automated analysis in a server’s robot.txt file (which allows server operators to put conditions on the use of downloaded materials); joining private groups, even where they are large – possibly using a pseudonym; persuading existing private group members to provide access; intercepting private communications;[123] and circumventing technical access control mechanisms (‘hacking’).

These are all significant interferences with individual rights, and therefore face a high bar in terms of the human rights framework. They should be carefully targeted, including in terms of the data collected, and have a reasonable likelihood of preventing significant harm, as well as being independently authorised and carefully overseen. Bartlett et al. suggest an individual surreptitiously joining a private online group is “analogous to an undercover agent joining

[121] Omand, Bartlett and Miller, note 19.

[122] *Ibid.*

[123] Jamie Bartlett, Carl Miller, Jeremy Crump and Lynne Middleton, *Policing in an Information Age*, London: Demos, 25 March 2013, p.10, at www.demos.co.uk/files/DEMOS_Policing_in_an_Information_Age_v1.pdf?1364295365

an offline group or organisation", while getting access via an existing group member would make the latter a "covert human intelligence source" in the UK legal framework.[124]

The use of encryption to protect data while stored and as it is transferred over the Internet makes it significantly more difficult to access for third parties – whether police or criminals. Internet companies are increasingly making use of encryption, partly as a response to consumer concerns following the Snowden revelations. Some politicians have proposed restrictions in response, with UK prime minister David Cameron asking in early 2015: "In extremis, it has been possible to read someone's letter, to listen to someone's call, to mobile communications ... are we going to allow a means of communications where it simply is not possible to do that? My answer to that question is: no, we must not".[125]

The EU Counter-Terrorism Coordinator has proposed that "The [European] Commission should be invited to explore rules obliging internet and telecommunications companies operating in the EU to provide under certain conditions as set out in the relevant national laws and in full compliance with fundamental rights access of the relevant national authorities to communications (i.e. share encryption keys)".[126]

There was an extended public debate throughout the 1990s, as encryption tools were originally developed, about police access to encrypted data. The broad conclusion reached then was that security gains to society from broader use of encryption outweigh the negative impact on police access to data for investigations. Attempts to find compromises that might enable police access, such as the use of weakened encryption algorithms or 'key escrow' systems, foundered

[124] Bartlett, Miller, Crump and Middleton, note 123.

[125] Nicholas Watt, Rowena Mason and Ian Traynor, "David Cameron pledges anti-terror law for internet after Paris attacks", the *Guardian*, 12 January 2015, at www.theguardian.com/uk-news/2015/jan/12/david-cameron-pledges-anti-terror-law-internet-paris-attacks-nick-clegg

[126] Council of the European Union, EU CTC input for the preparation of the informal meeting of Justice and Home Affairs Ministers in Riga on 29 January 2015, DS 1035/15, 17 January 2015, p.10.

on the difficulty of stopping such weaknesses being exploited by criminals (and the continuing availability of non-weakened tools to those that chose to use them). Computer security experts believe this decision remains correct, with a recent review concluding:

> In permitting continued use of communications devices employing encryption without law-enforcement backdoors, we are opting to trade one type of risk – inability to protect our communications from criminals and other eavesdroppers (e.g. nation-state spying) – for another – some inability to listen in (or immediately listen in) to terrorists' and criminals' communications. But despite law-enforcement's continued vehemence about the need for cryptographic backdoors, there's not an equivalence between the two types of risks. Our society's heavy reliance on electronic communications for everything from banking transactions to business communications makes securing electronic communications crucial – and the correct security choice.[127]

There are also sensitivities about state use of "hacking" tools to covertly access information, because they enable access to (and also changing of) potentially all of the data on a targeted system. In the most significant legal assessment to date of the compatibility of this practice with fundamental rights, the German Constitutional Court found that it was only justified "under strict conditions and when there is an imminent threat to the life, physical integrity or liberty of persons, or to the foundations of the state or the existence of mankind".[128] Evidence retrieved from a hacked system could also be challenged in any court proceedings.

[127] Susan Landau, "Finally ... Some Clear Talk on the Encryption Issue", *Lawfare*, 16 February 2015, at www.lawfareblog.com/2015/02/finally-some-clear-talk-on-the-encryption-issue/

[128] Didier Bigo, Sergio Carrera, Nicholas Hernanz, Julien Jeansboz, Joanna Parkin, Francesco Ragazzi and Amandine Scherrer, *National programmes for mass surveillance of personal data in EU Member States and their compatibility with EU law*, European Parliament, PE 493.032, October 2013, p.71.

'BIG DATA' ANALYTICS

Advances in computing and storage technology open up new opportunities for making use of the very large quantity of potentially extremist material available online. So-called 'big data' analytical techniques enable information to be extracted from very large data sets, which are characterised by high volume, velocity and variety.[129]

The accuracy of this type of analysis is constrained by the ambiguity of definitions of extremism; the importance of context in interpreting materials, especially free-form text; and the ease with which false and misleading material can be posted. These factors mean that analytics are more likely to be useful tools for investigators and researchers, rather than automated processes for accurately identifying all illegal – let alone extremist – content.[130]

Such techniques can be used, for example, to map flows of information through networks of supporters of particular extremist groups on social media. Carter, Maher and Neumann carried out such an analysis, finding that "a large number of foreign fighters receive their information about the [Syrian] conflict not from the official channels provided by their fighting groups, but through so-called disseminators – unaffiliated but broadly sympathetic individuals who can sometimes appear to offer moral and intellectual support to jihadist opposition groups". This study also found "new spiritual authorities who foreign fighters in Syria look to for inspiration and guidance... the two most prominent of these new spiritual authorities" based in the US and Australia.[131]

Another example of a social network analysis of social media by Berger and Strathearn "devised a scoring system to find out

[129] Mark Beyer, *Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data*, Gartner, 10 July 2011.

[130] Saltman and Russell, note 57.

[131] Joseph A. Carter, Shiraz Maher and Peter R. Neumann, *#Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks*, London: International Centre for the Study of Radicalisation, 2014, pp.1–2, at http://icsr.info/wp-content/uploads/2014/04/ICSR-Report-Greenbirds-Measuring-Importance-and-Infleunce-in-Syrian-Foreign-Fighter-Networks.pdf

which social media accounts within a specific extremist circle were most influential and most prone to be influenced (a tendency we called exposure)... Our starting data centred on followers of 12 American white nationalist/white supremacist 'seed' accounts on Twitter. We discovered that by limiting our analysis to interactions with this set, the metrics also identified the users who were highly engaged with extremist ideology".[132]

Online information still only forms part of extremism investigations, combined with significant amounts of information from traditional human intelligence sources, and there is still nervousness about the reliability of automated decision-making in such a complex area. A law enforcement officer told us: "A person always makes a decision about whether to investigate. The big problem is the volume of online material. Knowing what you need to be looking at tends to come from people not machines. Investigations are person or group specific, which doesn't lend itself to mass algorithms to work out who-to-who".

A SecDev Group study in Canada took a public health approach, looking for geographies and communities at risk of radicalisation. The researchers used analytics to undertake a very large pattern of life study, comparing the results with interviews as ground truth. While this proved very effective, it raised the question of what community interventions would be appropriate in identified areas. The researchers also found that analytics used to identify groups at risk could also be used almost seamlessly to identify individuals within those groups – meaning that there is no easy technology-based line that can be drawn between group and individual identification.[133] This raises a threshold problem – when will analysis of online material provide reasonable cause for suspicion justifying further investigation of an individual?[134]

[132] J.M. Berger & Bill Strathearn, *Who Matters Online: Measuring influence, evaluating content and countering violent extremism in online social networks*, London: International Centre for the Study of Radicalisation, March 2013, at http://icsr.info/wp-content/uploads/2013/03/ICSR_Berger-and-Strathearn.pdf

[133] The SecDev Group, note 84.

[134] Prabhu, note 61.

There is also the question of whether such automated profiling reduces discrimination, by avoiding conscious and unconscious stereotyping by human investigators of minority communities, or enables guilt by association.[135] Automated analysis may be able to avoid more obvious discriminatory profiling, so long as it focuses on attributes where evidence supports a link with violent extremism. As the UK Security Service found, this does not include ethnicity, religiosity, demography or mental illness.[136] But it is important to ensure profiling does not introduce proxies for such variables, or subtle but systematic discrimination that inadvertently perpetuates "the far more massive impacts of system-level biases and blind spots with regard to structural impediments that magnify the impact that disparities in starting position will have on subsequent opportunities".[137]

The SAPIENT project, which assessed the privacy impact of "smart surveillance" technologies through engagement with civil society and security stakeholders, recommended that "their operations, and their interactions with other elements, should equally be the object of a series of controls, including *ex-post* checks, to ensure that discrimination is not taking place".[138]

Finally, the use of 'big data' techniques raises the question of how much data should be collected and retained for analysis. The general human rights principle is that collection and retention are interferences with rights that must therefore be justifiable. This is more likely when higher-risk individuals are targeted, and/or more serious harms are likely to be prevented as a result. Speculative investigations, and analysis techniques that are unlikely to produce concrete results for investigating or preventing serious crimes, are less likely to be justifiable. Safeguards on the use and retention of collected data are also important.[139]

[135] Prabhu, note 61.

[136] UK Security Service, note 78.

[137] Oscar Gandy, "Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems", *Ethics and Information Technology 12*(1), 2010, p.34.

[138] SAPIENT consortium, note 23.

[139] Bartlett, Miller, Crump and Middleton, note 123.

RESEARCH

The previous sub-sections show how the diversity of environments within which extremist material is created and distributed online not only affects how useful this information is for law enforcement, but also requires a differential and flexible response to the legal and ethical issues raised by its use. Every online communication platform has different (and ever-evolving) technical functions and social norms, making the writing of hard-and-fast legislation or even best practice guidelines a significant challenge. Improving understanding of how extremist activity online is observed and analysed points a way forward, and academia and civil society groups have a powerful role to play in this process.

Every online communication platform has different (and ever-evolving) technical functions and social norms, making the writing of hard-and-fast legislation or even best practice guidelines a significant challenge. Improving understanding of how extremist activity online is observed and analysed points a way forward, and academia and civil society groups have a powerful role to play in this process.

Research conducted by academic and civil society groups has made significant contributions to wider understanding of online extremism. In their investigation of the use of the Internet for extremist activities, Hussain and Saltman found that the Internet is very rarely the initiator of radicalisation, but rather an incubator of extremist sympathies.[140] They show that it is unlikely for extremist material to simply be ‘chanced upon’, but rather that “the vast majority of those that visit extremist websites and consume the content enthusiastically are likely to have been heading in that direction”. Therefore, the Internet’s primary role is

[140] Ghaffar Hussain and Erin Marie Saltman, *Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter it*. London: Quilliam, May 2014.

one of mediation and facilitation, forming a bridge between the initial exposure to extremist narratives, and the perpetration of violent acts – both of which remain largely offline phenomena. Thus, Hussain and Saltman argue, removing objectionable (though not necessarily illegal) content "as a reaction to violent extremist acts is attacking a symptom, rather than dealing with the source, i.e. proliferation of extremist narratives in society".[141] Another prominent example of important research conducted in the civil society sphere is the *#Greenbirds* study by the International Centre for the Study of Radicalisation and Political Violence.[142] As noted above, the study cast light on how the Internet is used as a means of disseminating information and communicating across borders, thus helping to uncover the ground-breaking uses of digital media in the Syrian conflict and showing how these intertwine with the physical situation on the ground.

As well as yielding important substantive findings such as these, academic and civil society groups have also driven innovative methodological techniques to better understand the nature of online extremism. A prominent example of this is social network analysis. As Bartlett and Miller note, social network analysis is "at its root a sociological and mathematical discipline that pre-dates the Internet and social media".[143] The approach involves the mapping of social relationships as a means of gaining greater understanding of how networks operate. If used effectively, the utility of this approach for law enforcement – not least countering extremism – is obvious. Moreover, in cases where the information is available publically it is also potentially a less invasive technique – enhancing both its operational and ethical validity, since analysis can be done covertly.

A case study by Duijn and Klerks helps to illuminate how social network analysis is already being used in the disruption of

[141] *Ibid.*, p.61.

[142] Carter, Maher and Neumann, note 131.

[143] Jamie Bartlett and Carl Miller, *The State of the Art: A literature review of social media intelligence capabilities for counter-terrorism.* London: Demos, 2013, p.35.

crime in the Netherlands.[144] In the ‘Blackbird’ operation, Dutch law enforcement investigated a crime ring using social network analysis techniques, with key actors mapped and the most vulnerable parts of the network identified. Notably, in this investigation, the data analysed was obtained primarily through traditional means – for example via wiretaps and eyewitness statements – however, some data from online social media sites was also used to yield further information. Significantly, other research indeed suggests that criminal network dynamics often carry over to the Internet,[145] and network analysis purely using online data has found success, such as the identification of an illicit network of Russian drug users on the Russian social network Livejournal.[146]

Social network analysis therefore holds promise as an example of a technique adopted from earlier academic research traditions and adapted to the needs of modern law enforcement in particular contexts. Yet using online social media data to discover illicit networks of crime nonetheless requires a sensitive understanding of how communities engage on the Internet. Traditional metrics of social network analysis, such as ‘degree’ and ‘betweenness’, in some cases need to be updated to accommodate the specific dynamics of online networks. Bartlett and Miller point to a study in which the authors developed their own measures of influence within a network of white nationalist Twitter accounts, based on specific metrics of Twitter use such as retweets and replies.[147] This example points to the importance of understanding online extremist activities on their own terms and in relation to the data available: data from Twitter’s

[144] Paul A. C. Duijn and Peter P. H. M. Klerks, “Social Network Analysis Applied to Criminal Networks: Recent Developments in Dutch Law Enforcement”. In A. J. Masys (ed.), *Networks and Network Analysis for Defence and Security*. Switzerland: Springer, 2014.

[145] D. Décay-Hétu and C. Morselli, “Gang presence in social network sites”, *International Journal of Cyber Criminology* 5(2):876–890, 2011.

[146] L. J. Dijkstra, A.V. Yakushev, P.A.C. Duijn, A.V. Boukhanovsky and P.M.A. Sloot, “Inference of the Russian drug community from one of the largest social networks in the Russian Federation”, *ArXiv:1211.4783v2*, 2012.

[147] Bartlett and Miller, note 143, p.38.

The use of purely algorithmic techniques threatens to reduce the role of human decision-making, with both operational and ethical consequences.

API can only tell researchers so much. As such, Duijin and Klerks emphasise the importance of validating findings from online research by comparing results to other criminal intelligence sources as ground truth.[148]

The importance of validation becomes clearer in cases where the use of online data to predict human behaviour breaks down. For example, Google's Flu Trends service, which utilised search queries to estimate rates of influenza in the United States, was hailed as a breakthrough 'big data' approach to a public policy problem – that is, until it 'broke' in the winter of 2012–13, wildly over-estimating the prevalence of influenza amongst the population that season.[149] Such an example is disturbing enough in the case of public health, but the ethical consequences of wrongly predicting crime are even more unsettling.

Indeed, one need no longer rely on the fictional example of the Tom Cruise film *Minority Report* to reflect on the dangers of imperfect crime prediction: examples have emerged in real-world frontline policing. In 2014 it emerged that the Chicago Police Department was using predictive analytics to identify 'potential' criminals using geographic and social cues. The technical lead of the project, Miles Wernick, noted when the story came to light that the system identified people who, according to the algorithm, "clearly have a high likelihood of being involved in violence."[150] Yet, as noted above, the use of purely algorithmic techniques threatens to reduce the role of human decision-making, with both operational and ethical consequences. In the Chicago case,

[148] Duijn and Klerks, note 144, p.52.

[149] David Lazer, Ryan Kennedy, Gary King and Alessandro Vespignani, "The parable of Google Flu: traps in big data analysis", *Science*, 14 March 2014.

[150] Matt Stroud, "The minority report: Chicago's new police computer predicts crimes, but is it racist?" *The Verge*, 19 February 2014, at www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist

the consequences for the 'suspects' were relatively innocuous: some were visited by police and warned against committing crime. Yet it is not difficult to imagine the same technical approaches being deployed to counter of violent extremism, with the consequences for 'suspects' being more severe.

Other recent academic research points to the ethical and operational dangers not of inaccurate prediction, but for the potential manipulation of Internet users. Two academic studies conducted on Facebook illustrate this concern. In the first study by Bond et al, a large sample of American voters were exposed to different kinds of information on the day of an election.[151] Some saw a dialog box reminding them to vote, while others received an enhanced prompt featuring information about which of their friends had already voted. By using data from voting records as ground truth, the researchers were able to demonstrate a statistically significant increased likelihood to vote when users received the social information prompt. The second study, by Kramer et al, also investigated the effects of social contagion on Facebook, but in this instance on expressions of emotion.[152] The study divided a large sample of users into two groups: one group received a reduction in exposure to positive content in their news feed, while the other was exposed to less negative content. By analysing the emotional content of subsequent posts by the exposed users, the study detected the social transference of emotional states over Facebook: users exposed to positive (negative) content subsequently expressed positive (negative) emotions on the network.

[151] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle and James H. Fowler, "A 61-million-person experiment in social influence and political mobilisation", *Nature* 489, No. 7415, 2012, pp. 295–298.

[152] Adam D.I. Kramer, Jamie E. Guillory and Jeffrey T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks", *Proceedings of the National Academy of Sciences 111*, No. 24, 2014, pp.8788–8790.

These examples – particularly the latter, which prompted a media outcry and legal backlash[153] – highlight the danger of leaning too heavily on academic research for investigating human phenomena. The notion of intervention-based studies which manipulate the behaviour of research subjects is nothing new in the laboratory. Yet the notion of adopting these techniques for law enforcement – including the countering of extremism online – should be concerning due to the ethical dangers of overreaching in this area: people affected by the actions of law enforcement are criminal suspects, not consenting research subjects. Evidently, not every 'state of the art' methodological approach should be borrowed from academia for law enforcement purposes.

Therefore, the adoption of methodological techniques which emerge from academia need to be taken on a case-by-case basis in terms of their utility and legitimacy as tools for law enforcement. What is clearer, however, is that academics as well as civil society groups can and do offer distinct perspectives on the issue of online extremism, often in contrast to prevailing views in government and law enforcement. Work by the civil society group Quilliam, for example, takes a broad perspective on the government's counter-extremist Prevent strategy, serving to offer a broader range of policy options in terms of how extremist material should be tackled. The study by Saltman and Russell frames techniques like the blocking, censoring and filtering of objectionable material as 'negative measures', in contrast to 'positive measures' including the development of contrasting narratives. They argue that such measures are often counter-productive, since material will often reappear quickly and users will be driven to increasingly popular anonymous networks and the 'dark web'.[154] Moreover, they point out that it is not yet possible to unambiguously classify material is illegal (rather than

[153] Samuel Gibbs, "Privacy watchdog files complaint over Facebook emotion experiment", the *Guardian*, 4 July 2014, at www.theguardian.com/technology/2014/jul/04/privacy-watchdog-files-complaint-over-facebook-emotion-experiment

[154] Saltman and Russell, note 57, p.11.

Kneejerk responses to clamp down on social media use will stifle freedom of expression without usefully countering the threat of further atrocities.

merely objectionable) and therefore subject to removal; indeed, the same objectionable material has been repurposed for countering extremism by the US State Department, and can be used by law enforcement for better understanding the nature of online extremism.[155]

Similar arguments are made in the publication 'How to Beat the Media Mujahdeen' by the think tank Demos. The authors, Bartlett and Fisher, argue that the legally mandated taking down of illegal content should not be conflated with limiting the effectiveness of jihadist groups. As they point out – in concert with findings from Hussain and Saltman noted above – "there is no evidence that just watching online propaganda turns anyone into a terrorist". Their recommendation, instead, is to "use take-downs and account suspensions strategically, focussing on how the removal of accounts impacts on the ability of the network as a whole to operate".[156]

Finally, work by Phillip Howard, a professor in the Department of Communication at the University of Washington, has also fought back against prevailing arguments in favour of greater monitoring and blocking of online social media sites in light of terrorist atrocities. Howard points out that, having studied online extremism for several years, he has "found that whenever technology diffusion has been unfettered, radical Islam has dissolved", for example in the case of radical political parties moderating their message in the information-rich environment of the Internet.[157] On a behavioural level, too, Howard argues that positive messages tend to rise to the top on digital networks, such that "social media prevents negative

[155] *Ibid.*, p.8.

[156] Bartlett and Fisher, note 116.

[157] Philip N. Howard, *The Myth of Violent Extremism*, Yale Books Blog, February 6 2015, at http://yalebooksblog.co.uk/2015/02/06/myth-violent-online-extremism/

herding".[158] Drawing on this evidence, Howard suggests that kneejerk responses to clamp down on social media use will stifle freedom of expression without usefully countering the threat of further atrocities.

These perspectives help to demonstrate the crucial importance of academic and civil society groups in countering prevailing views of politicians, pundits and law enforcement in relation to the balance between liberty and security in the age of digital media. Yet as well as offering original research findings, innovative techniques and distinctive perspectives, many of these civil society groups also have a direct role to play at the frontline of efforts to counter extremism online. In many cases, civil society is far better placed than government to actively engage at a human level with those vulnerable to extremist narratives. Saltman and Russell for example suggest that "counter[extremist] speech initiatives should be civil-society led".[159] The government's role, in contrast, should be better circumscribed: government should offer transparency regarding the facts at hand, rather than "being seen to 'argue' against propaganda [which] may be perceived as the government lowering itself to the extremists' level".[160] Civil society, by contrast, is better placed to dispute and disrupt messages more directly.

Similarly, Bartlett and Fisher emphasise the importance of sharing what law enforcement investigations reveal about extremist actors online with those people tackling extremist narratives at the front line, since "in the end, it won't be governments … alone that play the decisive role in beating terrorist propaganda".[161] Phillip Howard also argues for "more direct dialogue between human rights groups, technology firms and government", pointing to the Global Network Initiative, an association that supports the establishing of shared goals and improved trust between civil society groups and governmental organisations.[162]

[158] *Ibid.*

[159] Saltman and Russell, note 57, p.2.

[160] *Ibid.*, p.9.

[161] Bartlett and Fisher, note 116.

[162] Howard, note 157.

Research by academia and civil society groups plays an indispensable role in understanding the web as an arena in which extremist activities take place. The various roles played by these groups – including not only providing original research findings and innovative methodological techniques, but also the voicing of distinctive perspectives and involvement at the front line of countering extremism – demonstrate the importance of maintaining and strengthening ties between the spheres of academia and civil society groups on one side and those of government and industry on the other.

ACTORS

IN THIS SECTION, we use an ‘ecology of games’ framework to examine the activities of the key actors interacting to produce policy outcomes in the area of policing online extremist material. This is a long-used approach in political science, which was further developed by Dutton et al. to examine freedom of expression and connection online for a 2011 UNESCO study. It emphasises the impact of cooperation and competition between different actors, each with separate but overlapping concerns, goals, and tactics, within an overall framework of rules. The overall policy process and outcomes evolve through the often-unanticipated interactions between stakeholders.[163]

These processes take place within broader public policy debates over Internet regulation, which add key constraints to counter-extremism policy options, as described by Dutton et al. in Table 1 (p.70).[164]

We examine here the activities of three main groups of actors shaping the outcomes of policing of online extremist materials: law enforcement and intelligence agencies; the Internet industry; and civil society groups and individuals. The European Commission has offered support to all of these groups “in their efforts to … keep illegal content from public access”.[165]

[163] William H. Dutton, Anna Dopatka, Michael Hills, Ginette Law and Victoria Nash, *Freedom of Connection, Freedom of Expression: The changing legal and Regulatory Ecology Shaping the Internet*, Paris: UNESCO Publishing, 2011, pp.23–24, at http://unesdoc.unesco.org/images/0019/001915/191594e.pdf

[164] *Ibid.*, p.24.

[165] European Commission, *Preventing Radicalisation to Terrorism and Violent Extremism: Strengthening the EU’s Response*, COM(2013) 941 final, January 2014, at http://ec.europa.eu/dgs/home-affairs/e-library/documents/policies/crisis-and-terrorism/radicalisation/docs/communication_on_preventing_radicalisation_and_violence_promoting_extremism_201301_en.pdf

Table 1: The Ecology of Freedom of Expression on the Internet

CATEGORIES	OBJECTIVES DEFINING CHOICES IN GAMES
Digital rights	Access – freedom of connection Freedom of expression Censorship Equality, e.g. media literacy and skills Freedom of information (FOI) Privacy and data protection
Industrial policy and regulation	Intellectual property rights (IPR): copyright IPR: patents Competition Technology-led industrial strategies ICT for development
User-centric	Child protection Decency: pornography Libel: defamation Hate speech Fraud
Internet-centric	Internet governance and regulation Domain names and numbers Standard setting: identity Net neutrality Licensing, regulation of ISPs
Security	Secrecy, confidentiality Security against malware, such as spam and viruses Counter-radicalisation National security, counter-terrorism

Source: Dutton et al. (2011)

LAW ENFORCEMENT AND INTELLIGENCE AGENCIES

Law enforcement and intelligence agencies aim to prevent and find evidence of serious criminality, and safeguard individual and community safety. They do this through investigations, citizen engagement, international cooperation, and requesting or requiring content to be taken down and blocked, and accounts be suspended.

Investigations and citizen engagement

According to our interviewees, law enforcement agencies carry out violent extremism-related investigations mainly in response to human intelligence (such as tip-offs). This is one reason why the maintenance of trust between police and communities is so vital.

Social media can be an effective channel for police to share information with the public to build trust, increasing transparency and providing reassurance about ongoing operations. A national example is Norway, where the Norwegian Board of Technology found that police stations have had great success using this mechanism. A notable example was the Twitter feed of the operational command in Oslo, which in 2014 was the fifth most-followed Twitter user in Norway.[166]

Social media can be an effective channel for police to share information with the public to build trust, increasing transparency and providing reassurance about ongoing operations.

The independent commission on the Breivik attacks also recommended that online monitoring should be prioritised, and that the Norwegian Police Security Service (responsible for interior intelligence and security in Norway) should not hold back on targeted monitoring for fear of being criticised for political surveillance. But provisions in proposed legislation in 2014 to allow 'big data' aggregate analysis were rejected following the Snowden revelations. McCarthy has noted the huge disparity in power and knowledge between individuals and state agencies observing them.[167]

A law enforcement official told us: "We work on a basis of a person coming to peripheral interest, then aggregate lots of sensor information to decide if it's worth taking a closer look. Prevention has a wider stance – where we try to go and find the unknown unknowns – but there is a close inter-relationship between investigations and prevention".

[166] Prabhu, note 61.

[167] McCarthy, note 89.

Law enforcement agencies therefore need tools to search and analyse publicly-accessible online materials, including specialised tools for less accessible materials, such as those being developed in the US Defense Advanced Research Project Agency's Memex programme. These include capabilities to identify common objects and people in different images, and a 'Dark Web'/Tor services web crawler.[168]

Police also need processes for authorised, targeted access to semi-private and private communications during investigations. Many countries require police to make such requests to Internet Service Providers, although recent French laws allow police with a judge's permission to have direct access to ISP records, when justified by operational matters from broad security and crime prevention prospects – mainly for investigations.[169] However, in 2008, the Bulgarian Supreme Administrative Court blocked remote Ministry of Interior and security service direct access to retained communications data without a court order as incompatible with the constitutional right to privacy.[170]

The Internet industry has sometimes been critical of government proposals to update these surveillance powers as new communications tools have become widely used. An industry interviewee told us: "a criticism often made of surveillance proposals is that law enforcement has become accustomed to getting data in one way from one type of company (e.g. phone bills), and simply want the same capability with new media. There is a similar issue with adjusting from policing physical demonstrations to social media activity. The conflation of child sexual abuse imagery and violent extremist material is very misleading. The political debate has become

[168] Larry Greenemeier, "Human Traffickers Caught on Hidden Internet", *Scientific American*, 8 February 2015, at www.scientificamerican.com/article/human-traffickers-caught-on-hidden-internet/

[169] Ducol, note 56.

[170] Access to Information Programme, *The Bulgarian Supreme Administrative Court (SAC) repealed a provision of the Data Retention in the Internet Regulation*, 12 December 2008, at www.aip-bg.org/documents/data_retention_campaign_11122008eng.htm

deliberately disingenuous, and betrays a lack of understanding of the technology. People in government that do understand the technology won't challenge the political narrative, and just accept it what's politicians want".

The police use of surveillance powers is more controversial when they are undertaking 'disruption' activities, trying to prevent organised groups from being able to commit potential offences, rather than investigating crimes. While 'predictive policing' is a popular concept for technology suppliers, as yet there is limited evidence of its effectiveness in this context – does it reduce, resolve or cause a threat?[171]

A civil society interviewee told us: "Police should not be doing automated analysis, patrolling online public spaces – that's mass surveillance. They have to do police work to know they have enough evidence that they should go and look at a particular website. It chills all our speech if we are all in a dragnet. Courts have been clear on the fact that if a human has been able to construct a machine to carry out a task, it doesn't make it any less intrusive; it makes it possibly more intrusive in levels of analysis".

"Police should not be doing automated analysis, patrolling online public spaces – that's mass surveillance. They have to do police work to know they have enough evidence that they should go and look at a particular website."

Civil society interviewee

In many European countries, intelligence agencies support police investigations into terrorism and serious crime using covert capabilities and data shared with other countries' agencies under secret agreements. They also act on tip-offs from informants, and find links to individuals that are 'subjects of interest'. A UK parliamentary inquiry stated that the UK Security Service is at any one time investigating several thousand subjects of interest; merely looking at extremist material is not enough justification for deeper

[171] McCarthy, note 89.

surveillance.[172] Intelligence agencies also conduct less individually targeted investigations using large-scale analysis of public and private data sources, including online materials, travel records, and financial records.[173]

European intelligence agency activity must be compatible with the European Convention on Human Rights, even when it is directed towards protecting national security. As the European Court of Human Rights held: "Powers of secret surveillance of citizens, characterising as they do the police state, are tolerable under the Convention only insofar as strictly necessary for safeguarding the democratic institutions".[174]

Valentin noted a series of Romanian Constitutional Court judgments on legal provisions concerning data retention, pre-paid telephony cards and identification of users of free WiFi, which were struck down because of lack of proper guarantees against unauthorised access. The Court rejected the provisions' distinction between law enforcement and intelligence agencies, and held that both should need warrants to access this data.[175]

International cooperation

Online activities very often include transnational communications between users, and with services hosted in a different country to users. This means that almost all online policing will raise "contested questions of jurisdiction based on the nationality of users, their physical location, the physical location of servers hosting or visited by software, the physical location of 'victims' (personal, corporate or state), and the physical location or

[172] Intelligence and Security Committee of UK Parliament, *Report on the intelligence relating to the murder of Fusilier Lee Rigby*, HC 795, 25 November 2014.

[173] Intelligence and Security Committee of UK Parliament, *Privacy and Security: A modern and transparent legal framework*, HC 1075, 12 March 2015.

[174] *Klass v Germany*, note 14, §42.

[175] Valentin, at http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

nationality of other involved parties, such as ISPs and search engine providers".[176]

An important element of online policing is linking the IP addresses associated with digital evidence to the end-user responsible for that content – even when the user is in a different country to an investigating officer. The EU Counter-Terrorism Coordinator has noted: "In the law enforcement and judicial context, cross-border information about owners of IP addresses can take very long to obtain, given the need to use [mutual legal assistance treaty] tools. The [European] Commission should be invited to consider ways to speed up the process. In the meantime, existing best practices in the Member States to deal with this issue could be collected and shared. Eurojust could facilitate this process…".[177]

Because so many major Internet services are based in the USA, the US legal framework is particularly important for this issue. The UK government has been negotiating with the US to enable faster cross-border provision of user data outside immediate 'threat to life' situations, and in 2014 updated UK law to claim extraterritorial jurisdiction over foreign communications providers. However, an industry interviewee told us: "Most companies [are] still holding the line that US companies [are] not subject to the jurisdiction of UK law.[178] US industry is lobbying for domestic legal reform reform, but still talking about using Mutual Legal Assistance Treaties (MLAT), as we need a solution that is globally scalable and robust. An industry coalition report[179] makes the point that it is up to governments to make that mechanism work. The danger is that industry is put in a difficult position through failure to resolve MLAT adequately. The US Electronic

[176] Omand, Bartlett and Miller, note 19.

[177] EU Counter-Terrorism Coordinator, note 126.

[178] This is confirmed in the *Report of the Interception of Communications Commissioner*, HC 1113, 12 March 2015, §§5.18–5.21, at www.iocco-uk.info/docs/IOCCO%20Report%20March%202015%20%28Web%29.pdf

[179] Andrew K. Woods, *Data Beyond Borders: Mutual Legal Assistance in the Internet Age*, Washington DC: Global Network Initiative, January 2015, at https://globalnetworkinitiative.org/sites/default/files/GNI%20MLAT%20Report.pdf

Communications Protection Act, Stored Communication Act, Wiretap Act and consent decrees from the FCC set a much higher threshold for companies to make content than metadata disclosures to law enforcement agencies, but that is the case in most jurisdictions".

An important prerequisite for greater international cooperation is harmonisation between national standards for state access to user data. An industry interviewee told us: "The UK government is not going to allow UK interception warrants to be scrutinised by US courts because of the different US standard of scrutiny. But there are also important differences between evidence and intelligence requests – these are under different legal frameworks but also have different operational requirements. US-UK negotiations seem to be going well, and the US administration has increased funding to MLAT, so diplomatic progress can be made. The challenge is whether political pressure for action is too great to allow the diplomatic process to run its course".

As well as MLATs, regional cybercrime treaties – such as the Council of Europe's Budapest Convention on Cybercrime – include provisions on international cooperation. There are also formal cooperation institutions such as Europol and Interpol as well as informal links between countries. A law enforcement officer told us: "We can engage with international partners in lots of ways e.g. via the government's foreign service. It's a question of the effectiveness of our networks and the politics of whichever country you're trying to cooperate with. Sometimes going through a third party (such as Europol) makes it even more complicated". An industry interviewee added: "governments are getting the message that these issues can't be solved nation by nation, there has to be an international response. The White House is leading, saying you can't have hotchpotch laws, and the EU Council has referenced an Interpol-style model. This is a positive trend for industry and government, which will pull the debate back to the centre. Given work on MLAT reform this is likely to become a legally-based framework, with some elements of voluntary cooperation".

Brown et al. have analysed the processes by which an initially small group of democratic governments could develop an international agreement on human rights standards for foreign surveillance, dealing with the public disquiet resulting from the revelations by Edward Snowden of extensive online surveillance by North American and European governments. They analyse the four main reform proposals so far at the national or European level relating to such surveillance – from industry, civil society, a US review commission, and the European Parliament – and describe how new international protections can be put in place to ensure such surveillance is necessary and proportionate.[180]

A further obstacle to cooperation is the sovereignty concern of states over counter-terrorism. An EU official told us: "When it comes to the counter-terrorism mandate of Europol, there is very little willingness of member states to contribute information because they don't really trust them, since there are other member states with whom they have not such great links. No really sensitive intelligence is being forwarded to Europol in the counter-terrorism context. This is quite different compared to other transnational crime issues, such as combating child abuse images, where there is a better level of cooperation. This is mainly about nationalist politics, but agencies that would want to share this type of information also have a very strict need-to-know thought process and don't see the added value to give this information to a multilateral body, which doesn't have any executive powers to collect intelligence itself, so it all depends on intelligence fed by the member states. Many people think if they really need something they can just act bilaterally, which also allows them to bypass the Europol data protection framework, which is very robust and a lot of agencies still see it as a lot of unnecessary red tape".

Taking down content and accounts

A civil society interviewee told us: "states feel they have a responsibility to protect their citizens, and believe they should remove extremist material, and Internet companies should do something about it. Citizens, because they may not see it, may well ignore it,

[180] Ian Brown, Morton Halperin, Ben Hayes, Ben Scott and Mathias Vermuelen, Oxford Internet Instititute discussion paper, January 2015.

and yet the state has to try and protect everyone, whereas citizens are ultimately looking out for themselves. Extremist material will have an impact on a tiny minority of people who wish to do harm, with the broader public saying, for the majority of us it doesn't have any impact, takedowns look like an affront to freedom of speech and to an open and neutral web".

The EU Council has agreed that "adequate measures [should] be taken to detect and remove internet content promoting terrorism or extremism, including through greater cooperation between public authorities and the private sector at EU level, working with Europol to establish internet referral capabilities".[181]

These "referral capabilities" are based on the UK's Counter-Terrorism Internet Referral Unit (CTIRU), created in 2010 to "remove unlawful terrorist material content from the Internet". Online extremist material very frequently breaches online services' terms and conditions. CTIRU therefore refers online material they assess to be unlawful under sections 1–3 of the UK's Terrorism Act 2006 to non-UK based service providers for assessment and possible taking down on a voluntary basis. (Such material would be illegal for UK-hosted services to disseminate "recklessly" or following notice).

Since February 2010, CTIRU "has taken down 72,000 individual items" of unlawful terrorist-related content,[182] but there is a "lack of clarity over the content and context of the material that has been removed and no further information about how many of these take-downs are duplicates or re-posts".[183]

The EU Counter-Terrorism Coordinator has called for Member States to "consider establishing similar units to the UK CTIRU and replicate relationships with the main social media companies to refer terrorist and extremist content which breaches the platforms' own

[181] Statement by the members of the EU Council following the Informal meeting of the Heads of State or Government, CL15–030EN, 12 February 2015.

[182] James Brokenshire MP, House of Commons Debates, *Hansard* col. 332, 21 January 2015.

[183] Saltman and Russell, note 57, p.5.

terms and conditions (and not necessarily national legislation)".[184] An EU official told us: "there won't be any legal initiative as far as I know – the legal services in most Member States already figured out that the impact on freedom of speech would be too big and set a bad precedent. But that doesn't rule out memoranda of understanding, codes of conduct or other soft law initiatives of course".

An EU official told us it is important that such referral units only attempt to have content taken down that is illegal, since "social media platform Terms of Service define certain types of unwanted and undesired content, which are assessed in a completely different type of balancing exercise to judicial processes. Content is sent for assessment to people earning very low wages in the developing world. We don't even know the specific detailed rules in place, so to demand from public law enforcement authorities to enforce these norms as opposed to organically developed human rights rules is a completely different ball game".

"If every single ISIS supporter disappeared from Twitter tomorrow, it would represent a staggering loss of intelligence – assuming that intelligence is in fact being mined effectively by someone somewhere. However, many thousands of accounts can likely be removed from the ecosystem without having a dramatic negative impact on the potential intelligence yield."

Berger and Morgan

There is a tension between taking down content and accounts from online services, with the intelligence value of allowing related activity to be monitored. An industry interviewee told us: "The biggest issue will be government-to-government, since the preferred option could be different for the hosting jurisdiction and foreign governments. The biggest security risk is returning foreign fighters, and there's a value in those people using public platforms since it gives you intelligence".

[184] Council of the European Union, note 126.

Regarding ISIS-related Twitter account suspensions, Berger and Morgan noted: "the challenge is to sufficiently degrade the performance of the network to make a difference without driving the less visible and more valuable ISIS supporters out of the social network in large numbers. If every single ISIS supporter disappeared from Twitter tomorrow, it would represent a staggering loss of intelligence – assuming that intelligence is in fact being mined effectively by someone somewhere. However, many thousands of accounts can likely be removed from the ecosystem without having a dramatic negative impact on the potential intelligence yield".[185]

INTERNET INDUSTRY

Internet companies aim to provide communications services to their customers, in a socially responsible and trustworthy way, including by taking down content that is illegal or breaches their terms and conditions, and when appropriate providing user data to security agencies.

Takedowns, blocking and account suspensions

Internet companies such as web hosts and social media platforms already proactively remove large volumes of material from their services that is in breach of their terms and conditions, partly based on notification from users and also government agencies such as CTIRU, which have "priority flagging systems" or "specific reporting streams". Many providers also have internal teams that search out and take down such content, "often… more than the amount of content taken down through government flagged content".[186]

Alongside social media, Istvan Janto-Petnehazi has noted the relevance of user-generated content on traditional media sites such as national newspapers, given their sometimes very large readerships. He analysed the use of Romanian newspapers' online comment

[185] Berger and Morgan, note 86.

[186] Saltman and Russell, note 57, p.5.

sections to propagate hate speech to a general audience, and how effective site usage policies could be in reducing this.[187]

Some states have encouraged Internet access providers to block access to specific, identified violent extremist material as protection for children and vulnerable adults, including those with mental illnesses. This has become a bigger issue as children increasingly use mobile Internet access with less adult supervision.

Because of the contextual analysis needed to identify violent extremist content, more automated blocking would be "technically and legally contentious, and practically impossible to implement".[188] Takedowns and blocking can also be ineffective as content is "likely to reappear in another format", a phenomenon frequently observed with ISIS-related content.[189]

In the UK, material reported by CTIRU is also blocked by Internet Service Providers for the "public estate" (such as schools, colleges, public libraries, and immigration removal centres). The Home Office (interior ministry) distributes a list of the web addresses of this content to filtering software companies, whose software is used by the "public estate" providers, but is likely also being included in other filtering lists, such as the "family friendly" default filters now being applied by many UK ISPs to all of the customers that have not opted-out of this. This is all done on a 'voluntary' basis – the Home Office signs a licence deed with participating companies, which contains a confidentiality clause.[190]

No impact assessment has been made public, so it is not clear how issues such as human rights compatibility and liability for blocking mistakes are dealt with. Blocked sites are not notified before or after the fact, and have no appeal mechanism. A user attempting to access the material knows the site is blocked, but not why.[191]

[187] Istvan Janto-Petnehazi, at http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

[188] Saltman and Russell, note 57, p.7.

[189] Saltman and Russell, note 57, p.10.

[190] McIntyre, note 111.

[191] *Ibid.*

An industry interviewee asked: “If you have network-level filtering of supposedly extremist material does that breach interception laws? Will ISPs be censoring material that is political speech, and may be non-violent? Some ISPs are now putting filters on default for users. There is not much transparency on how those filtering lists are put together, and many examples of how blocking is restricting access to all kinds of content. It puts ISPs in a difficult situation by changing the nature of ‘family-friendly’ filters, and reflects wider policy failure since people don’t become radicalised by accidentally stumbling across something on the Internet. Where can these issues be discussed politically? Industry is taking practical steps to remove content when brought to its attention. No company has ever said it won’t take CTIRU reports. The danger is the policy debate becomes coloured by the notion you can short-circuit hard legal questions by applying political pressure to industry. There needs to be more work to address these hard challenges”.

Putting too much responsibility on Internet companies to police their networks would be likely to have a very significant impact on freedom of expression.

Berger and Morgan have suggested that “it is highly likely that Twitter could – if it so chose – substantially deny the use of its service to ISIS supporters, reducing their ranks to as few as a couple hundred hyper-committed supporters with negligible influence. For many reasons – including issues associated with establishing a broad precedent on political speech and the practical intelligence [implications] – we do not recommend this approach. However, it remains theoretically possible… It is also possible to fine tune the current suspensions efforts to further limit Twitter’s utility to ISIS, without completely eliminating the group’s presence. For instance, given the large number of small accounts in the system, we believe it would be possible to design metrics … that could be used to dismantle the network by separating these small accounts into ever smaller clusters of users, and disrupting the flow of information among them”.[192]

[192] Berger and Morgan, note 86.

Putting too much responsibility on Internet companies to police their networks would be likely to have a very significant impact on freedom of expression, and according to Philip Howard "would undermine the development of new social networking applications and dampen the many benefits that come from vibrant digital media. There are plenty of crackpots, angry people, and bad ideas on social media. But in practice, crackpots get marginalised and bad ideas are dismissed. It is far from clear that Twitter and Facebook amplify extremist voices to any advantage, since civic and moderate voices get amplified too".[193]

The former UN Special Rapporteur on Freedom of Expression recommended that Internet intermediaries should "only implement restrictions to [freedom of expression and privacy] after judicial intervention; be transparent to the user involved about measures taken, and where applicable to the wider public; provide, if possible, forewarning to users before the implementation of restrictive measures; and minimise the impact of restrictions strictly to the content involved. Finally, there must be effective remedies for affected users, including the possibility of appeal through the procedures provided by the intermediary and by a competent judicial authority".

Care is needed that measures taken by social media platforms cooperating with law enforcement agencies does not encourage extremists to migrate to smaller, less cooperative platforms. Saltman and Russell have found that "as certain social media platforms are seen as less accessible for extremist and terrorist networks, these targeted actors are seen moving to less restrictive and more anonymous platforms, such as Vkontakte, Kik and Snapchat".[194]

A law enforcement officer told us: "We have to do something to get rid of horrific stuff like beheadings – the victims' families want us to get rid of it. But is trying to take down and block extremist material effective for restricting the extremist use of the Internet

[193] Philip N. Howard, "The Myth of Violent Online Extremism", *Yale University Press blog*, 6 February 2015, at http://yalebooksblog.co.uk/2015/02/06/myth-violent-online-extremism/

[194] Saltman and Russell, note 57.

and preventing radicalisation/minimising impact due to public messaging? No. Should we be doing it? Yes, probably. It's a crude tool that has some value. Providers think it's a good thing, since it takes responsibility off them. When content doesn't fall within offences, then it's very difficult, and a lot of what is publicly agreed is extremist material doesn't have a framework around it. Beheading videos and how you blow stuff up videos – you're too late by that stage. Individuals such as the three London teenagers that travelled together to Syria in early 2015 have developed a relationship with ISIS members over a long period of time. These are different things. The latter is far harder to deal with. There's no offence of communicating with anyone. It's a safeguarding issue – they're kids, not the most dangerous people. If we're trying to protect the public, we're more worried about someone who's going out fighting, building capability, then coming back. But we should stop them going out there because they will probably be killed".

Provision of user data

Internet companies that provide services 'over the top' of Internet Service Providers – such as webmail and social media platforms – have come under pressure to voluntarily provide user data to law enforcement agencies outside their own jurisdictions, because Mutual Legal Assistance Treaty applications can take months, and because the increasing use of encryption means that this data is increasingly difficult to directly intercept. An industry interviewee claimed that such government requests, as well as to use encryption that can be broken by government agencies, are an attempt by "policymakers to try and get back to where we were three years ago. Companies are encrypting to build user trust, not to anger governments. Most social media platforms have users in conflict zones [at much greater risk]. They can't selectively encrypt".

Government requests for direct (or 'bulk') access to Internet company databases have been strongly resisted. An industry interviewee told us: "So much data is openly available that you shouldn't need bulk access to do target discovery. Industry provision of more sophisticated querying capability is not even on the table. In Silicon

Valley any perception of a government ability to directly query user data is no-go, as is a model that allows queries to be run across ISP networks, as that removes the ability of a company to control access to its data. It would not be transparent or subject to scrutiny. Some governments are being more open about providing more context about requests, and on that basis companies then might be willing to do more data analysis to answer requests. This gives companies confidence they are not infringing on legal liability and rights of users, and also enables companies to play a role in data minimisation, only providing data that's absolutely necessary, not everything government wants".

McIntyre has described the problems that could be caused by voluntary reporting by Internet companies of suspected extremism, as suggested by the UK Parliament's Intelligence and Security Committee (ISC),[195] which could lead to a flood of false positive reports, and require very significant resources for analysis.[196] An industry interviewee told us: "There is a fundamental question whether platforms can hand over data not set out in law (e.g. without a warrant). There is no legal analysis in the ISC report of what they suggested. Proactive notification without an absolutely clear legal mandate is huge. The broader policy issue is that the scale of social networks is huge, you're expecting the social media staffer to decide if something is a genuine threat – but they are not an intelligence analyst. The only way that model works is if you merge the intelligence services and reporting functions of social media companies. Nor did the ISC understand the volume of information that could be provided under their model. If they had had a more formal discussion with experts and industry, they could have better reflected the technical, legal and ethical challenges. The volume would swamp intelligence agencies (imagine watching all the video), damage user trust, and encourage extremists to use harder-to-find platforms. As soon as you put a sticker on a platform saying we do proactive reporting,

[195] Intelligence and Security Committee of UK Parliament, note 3.

[196] McIntyre, note 111.

terrorists will leave, and non-democracies will use it for their own purposes. Nor is the idea supported by former intelligence officials".

An industry interviewee told us: "When governments are responding to specific incidents you get a lot more rhetorical discussion of policy rather than specific problem-solving. It isn't that industry isn't cooperating and discussing issues, but different companies have different interests. For example, encryption has very different implications for different countries. The portrayal of Silicon Valley as a technological libertarian cabal who don't care about terrorism and serious crime is pretty offensive. It is not true of day-to-day operations between various national law enforcement agencies and companies who take their obligations seriously".

McCarthy has noted that the rise of 'big data' means that many more companies will in future gather user information that may be of interest to law enforcement agencies, and will have more diverse interests than Internet companies. There will also be a continued growth in companies offering analytics services.

A law enforcement officer told us they can "see scenarios where it is useful to bring in outside analytical help if appropriate and proportionate. We want to be innovative, creative and collaborative where we can. But I don't think we will have machinery that just churns intelligence out, because it comes down to somebody reading it and applying knowledge and understanding and expertise to enrich that data and make a decision. If somebody has something to our benefit why wouldn't we use it, but we have finite resources. You could lose yourself in a prioritisation nightmare, like all the calls to anti-terrorism hotlines after a major incident, where people are panicking, thinking they have to, for example, tell someone a neighbour has bought fertiliser".

CIVIL SOCIETY AND INDIVIDUALS

Civil society groups and individuals take political and other action in pursuit of a wide range of goals – most relevant in this area being sharing of relevant information with public and private sector organisations to aid in investigations and safeguarding; promoting counter-extremist narratives online; and campaigning for human rights, community protection and cohesion.

McCarthy and others have noted the different ways in which citizens can work collaboratively online to support law enforcement, and "passive bystanders can become active citizens".[197] They can engage via social media, building trust relationships and acting as "citizen journalists, providing and relaying information from the ground", as well as reporting violent extremist content using specific links and buttons to social media platforms.[198]

Some civil society groups, such as the Southern Poverty Law Center and Anti-Defamation League (ADL), put significant work into identifying and reporting hate speech. ADL has also produced a "Cyber-Safety Action Guide" to help individuals report online hate speech to a range of Internet companies.[199] Andrea Cerase has suggested further mechanisms to improve the identification and reporting of hate speech to ISPs, social media platforms and law enforcement agencies.[200]

One civil society interviewee felt care should be taken with organised groups identifying extremist online speech, saying: "Individuals should always be able to report something if they're concerned. The problem is when those groups come together and do that in a more systematic way and end up taking on a law enforcement function, that's where we've got to be very careful. It can attract money and funding and then they are taking on the role of a state, and we should have better regulation".

User reporting tools can become politicised, as for example seen in attempts by supporters of Syrian president Bashar al-Assad to spuriously report and hence have anti-regime materials removed from

[197] Omand, Bartlett and Miller, note 19, p.24.

[198] *Ibid.*

[199] See www.adl.org/combating-hate/cyber-safety/c/cyber-safety-action-guide.html

[200] Andrea Cerase, At http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

social media platforms,[201] and the activities of the Vietnamese police to reportedly "coax Facebook into shutting down at least 30 pages of prominent activists and independent journalists".[202] Countering such tactics can be resource-intensive for Internet companies, requiring careful consideration of reports and processes such as user warnings, guidance, and avenues for appeal if content is removed.[203]

Citizens can respond to police requests that explicitly 'crowd-source' information – although a recent counter-terrorism investigation that used this tactic, following the Boston marathon bombing, illustrated the problem of a flood of false positives that can result. Citizens can also choose to act as "active sensors", contributing data by running smartphone apps.[204]

Offline monitoring, by families and communities, can play an important role in monitoring online extremist material. A civil society interviewee told us: "An example is a schoolchild telling their mum about a friend sharing extremist content online, who then tells the school, which notifies a local police officer. No reason why the police need to spend resource and capability in monitoring that individual, a lot of the time they can visit the person, send a cease and desist letter, quite a nice little flag from the state to say you should stop this, and likewise monitoring may be going round to this child's family and saying to them, we have seen your child look at so-and-so, we are concerned, we would like you to do the monitoring. It's not just the state, it's the family. Then you would need the family's permission to look at the computer. The child has done nothing illegal, the police

[201] Michael Pizzi, "The Syrian Opposition Is Disappearing From Facebook", *The Atlantic*, 4 February 2014, at www.theatlantic.com/international/archive/2014/02/the-syrian-opposition-is-disappearing-from-facebook/283562/

[202] Doan Trang, 16 July 2014, at www.facebook.com/longDobby/posts/10202590412946131

[203] Jillian York, "With Facebook's 'Reporting Guide,' A Step in the Right Direction", *TechPresident*, 27 July 2012, at http://techpresident.com/news/22478/op-ed-facebooks-reporting-guide-step-right-direction

[204] McCarthy, at http://voxpol.eu/events/workshop-on-the-ethics-and-politics-of-online-monitoring-of-violent-extremism/

are simply telling the family the kid may be potentially vulnerable, the content on their machine is of an extreme nature and you might want to do something about it. This is the same way police often deal with drugs and alcohol. That is monitoring but in a different way. Police don't have the time and resources to do it themselves anyway. A lot of the time the leads are going to be from human sources i.e. a member of an education establishment, health worker, housing estate manager".

Academic researchers and civil society groups can be better-positioned than government agencies to undertake broad online research into extremist movements, following the research ethics approach described in *Research* and focusing on trends and processes linked to violent extremism, hence reducing concerns about social surveillance. This understanding then helps both government and civil society actors to better address these processes.[205]

Counter-speech

Civil society groups are well-positioned to undertake counter-speech initiatives, which challenge violent extremist ideology in online discussions and other spaces. While governments can be effective at explaining policies and positions, civil society has specific expertise in addressing ideological claims; credibility with relevant communities; and the independence from government to reduce accusations of political propagandisation. It can better target specific vulnerable audiences and anti-authority vulnerable individuals, and track viewership trends and the impact of specific messages, which is not possible with 'negative' measures such as taking down and blocking content.[206]

Civil society groups have the opportunity to be more nimble in experimenting with different online messaging strategies. An industry interview told us: "Governments think ISIS is amazing at social media, but in reality governments just suck at it. ISIS is just operating

[205] Saltman and Russell, note 57.

[206] *Ibid.*

"Governments think ISIS is amazing at social media, but in reality governments just suck at it. ISIS is just operating in the same way as teens and digital natives."

Industry interview

in the same way as teens and digital natives".

EU Member States and the European Commission have supported civil society groups in counter-messaging projects[207] – although groups need to be careful to take steps to protect their independence when accepting such funding. This has less political import when unrestricted grants, training and other support are provided by industry, as for example Facebook, Google and Twitter have done, and has the benefit for industry of improving the quality of their platforms and creating a hostile environment for violent extremists.[208] An industry interviewee told us: "This shouldn't be any different to other media. In any forum you need moderate voices". Funding transparency remains vital, and an industry interviewee told us that "large American Internet companies are not universally popular either".

Equally important is the ability of all Internet users to critically evaluate political and ideological messages they encounter online. The EU Counter-Terrorist Coordinator has suggested this can be increased through enhanced Internet safety education in schools, with Sweden's critical thinking skills school programme as an example of good practice.[209]

A civil society interviewee told us that a risk with counter-speech initiatives is that they end up giving more publicity to extremist messages than would otherwise have occurred.

[207] European Commission, *Preventing Radicalisation to Terrorism and Violent Extremism: Strengthening the EU's Response*, COM(2013) 941 final, January 2014, at http://ec.europa.eu/dgs/home-affairs/e-library/documents/policies/crisis-and-terrorism/radicalisation/docs/communication_on_preventing_radicalisation_and_violence_promoting_extremism_201301_en.pdf

[208] Saltman and Russell, note 57.

[209] EU Counter-Terrorism Coordinator, note 126.

CONCLUSION

BROAD INTERNATIONAL CONCERNS about online violent extremist material have led states, industry, and civil society groups to take action to monitor and block access to this material. This has implications for a range of human rights protected under international law: most obviously privacy and freedom of expression; but also freedom of association and assembly, and of thought, conscience and religion. It raises difficult ethical and political questions about definitions of extremist material; actions that should be taken in relation to public, semi-private, and private material; and the responsibilities of these different actors.

Monitoring online violent extremist material in aggregate – to understand broad trends, and the evolving tactics and ideologies of extremist groups – can provide extremely useful information to states, which are responsible for protecting peoples' lives and safety against violent extremist acts. When lawful, necessary and proportionate, monitoring can be used to investigate individuals and groups suspected of planning or committing serious crimes.

New tools are needed for public safety policing alongside traditional targeted surveillance. But not all online content can be treated as actionable intelligence, and adding more 'hay' to 'haystacks' of available data will not necessarily help investigations. A law enforcement officer told us: "You can harvest all the information you want, but you have to have someone looking at it all. You'd need warehouses and warehouses full of people doing nothing else. Not even the Americans or Chinese have those". An example of these constraints is the case of Aqsa Mahmood, a Scottish recruit to ISIS suspected of recruiting three London teenagers, initially via social media, and encouraging them to fly to Turkey and then travel on to Syria to join the group. The teenagers' families have repeatedly asked why investigators had not detected this contact and warned them.[210]

Despite such concerns, and especially following the Snowden revelations of large-scale government Internet surveillance, more

[210] Ashley Fantz and Atika Shubert, "From Scottish teen to ISIS bride and recruiter: the Aqsa Mahmood story", *CNN*, 24 February 2015, at http://edition.cnn.com/2015/02/23/world/scottish-teen-isis-recruiter/index.html

than one of our interviewees referred to a “crisis of trust” that has developed between governments and publics over online monitoring, which will need significant further efforts to address. A law enforcement officer told us: “If policing is by consent, are police getting a clear message about what the public want them to do – broad bottom-up feedback? People don’t care about the 999 people we stop from committing a crime – the focus is on the one that succeeded. Public perception tends to be around failure not success. Post-event people always want the perfect system. Snowden means people must be able to see everything the state is doing”.

Particularly relevant in addressing this “crisis of trust” in the EU context is the reform of the data protection framework, both broadly (in the proposed Regulation[211]) and specifically for law enforcement (the proposed Directive[212]), if an EU institution such as Europol is asked to monitor online extremist content, and to coordinate sharing this and related criminal intelligence information between Member States.

An EU official told us such a duty would have important resourcing implications for Europol, which has been doing a “lot of extra work on their cybercrime centre, with not enough people and funds, so an enhanced Check The Web initiative would not fit in this context of budgetary restraint”. They also noted that given the importance of public-private partnerships between states and industry for dealing with online extremist material, data protection reforms must “avoid gaps in protection, protecting privacy when profiling is taking place”, and “ensure protection is not lost through use of ‘non-personal’ data”.

Alongside greater transparency, effective independent authorisation and oversight can ensure that safeguards for use of intrusive

[211] European Commission, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final, 25 January 2012.

[212] European Commission, Proposal for a Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data, COM(2012) 10 final, 25 January 2012.

monitoring techniques are followed. Care is needed that such investigations do not cause or reinforce overt, or less obvious, discrimination. Ongoing legislative scrutiny of state monitoring powers, with stronger attention to their impact on human rights, could also make better use of expert institutions (which in the EU includes the European Data Protection Supervisor and Fundamental Rights Agency), and systematically include those affected.[213]

The European Parliament has voiced a longer-term concern about large-scale online monitoring, concerning "the establishment of a fully fledged preventive state, changing the established paradigm of criminal law in democratic societies, promoting instead a mix of law enforcement and intelligence activities with blurred legal safeguards, often not in line with democratic checks and balances and fundamental rights, especially the presumption of innocence; recalls in that regard the decision of the German Federal Constitutional Court on the prohibition of the use of preventive dragnets ('präventive Rasterfahndung') unless there is proof of a concrete danger to other high-ranking legally protected rights, whereby a general threat situation or international tensions do not suffice to justify such measures".[214]

Given the core human rights test of necessity for interferences with privacy and other rights, the broad range of policy options that do not involve online monitoring targeting individuals is critical. As Saltman and Russell noted, these include "civil society action, engagement with extremist ideologies and narratives, development and dissemination of counter-narratives, and addressing the grievances perceived by those vulnerable to or experiencing radicalisation",[215] as well as "digital literacy and critical consumption education" and "addressing the grievances perceived by those vulnerable to extremism".[216]

[213] SECILE Consortium, note 13.

[214] European Parliament, note 16, §10.

[215] Saltman and Russell, note 57, p.11.

[216] *Ibid.*

The VOX-Pol Network of Excellence (NoE) is a European Union Framework Programme 7 (FP7)-funded academic research network focused on researching the prevalence, contours, functions, and impacts of Violent Online Political Extremism and responses to it.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 312827

Email info@voxpol.eu
Twitter @VOX_Pol
www.voxpol.eu