UNIVERSITY OF CAMBRIDGE

Bennett Institute for Public Policy Cambridge

POLICY BRIEF

The Online Safety Bill

by Ross Anderson and Sam Gilbert
with a preface from Diane Coyle

October 2022

Policy Brief - The Online Safety Bill

1. Preface

The UK's Online Safety Bill is continuing its slow progress through parliamentary procedures. But even as the Government of new Prime Minister Liz Truss has said it will continue with the legislation, some of its contents continue to be contentious. With a further round of debate in the House of Commons and then the House of Lords due to take place in the months ahead, this timely policy brief sets out some of the key issues. While there is broad agreement that some action is needed to address online harms, defining these is not straightforward. Nor is devising effective means of implementation and enforcement of some aspects of the Bill as it stands. With a complex piece of legislation that is extremely broad in its scope, the chances of unintended consequences are high — unless great care is taken at this stage of shaping the legislative detail.

The aim of this Policy Brief by Ross Anderson, Professor of Security Engineering at the Universities of Cambridge and Edinburgh, and Bennett Institute Research Affiliate Sam Gilbert, is to shed light on some important aspects of the legislation, ahead of the next stages of parliamentary debate. Although focused on the current UK policy decision, the issues raised have far broader implications, as many countries are considering how best governments can ensure the digital environment serves their society in the light of widespread concerns about the consequences of misinformation, polarisation, harms to young people, terrorist finance, surveillance, and other issues that have been part of the public debate. The concerns and the dilemmas raised are almost universal, while the planned UK legislation is among the earliest and most comprehensive.

One of the key conclusions of this Policy Brief is that while tech platforms need to deliver on their responsibilities, given their undoubted scope and power, they alone cannot fix deep-seated social problems, although they can exacerbate them. But neither should online harms be an excuse for the government piggybacking on the power of tech companies to do things that would not be compatible with democratic accountability offline. Moreover, some harms need offline interventions or may be better addressed offline; for example, tackling the abuse of children or vulnerable people requires the involvement of local police, social services, and schools, which all need to have a legal framework enabling them to act and adequate resources with which to do so.

There are certainly problems that need to be tackled; doing so effectively requires clarity about the nature of problems, the technologies, and the appropriate responsibilities and accountability. This policy brief offers such clarity about some of the contentious areas of the Online Safety Bill. The authors support a duty of care for important digital platforms but warn strongly against relying on online actions alone to guard against all the harms that have been discussed in the debate, and underline the importance of offline measures too.

Diane Coyle

2. Introduction

Public opinion with respect to a small number of tech platforms, including Facebook, Instagram, WhatsApp, Google, YouTube, Twitter and TikTok, shapes the context for the Online Safety Bill. Rightly or wrongly, many people believe these platforms largely constitute the web. Having previously been viewed as neutral or even benign, a series of scandals has raised questions about these major platforms' responsibility for serious harms that result from the services they provide.

In some cases, it is the direct abuse of the services that has led to harm – for example, the "groups" features of Facebook and WhatsApp were deliberately used to incite and co-ordinate the ethnic cleansing of the Rohingya people in Myanmar.[1] In other cases, the harms are more akin to externalities or unintended side-effects – for example, the availability of graphic images of self-harm on Instagram and Pinterest prompted 14 year old Molly Russell to take her own life.[2] Platform owners' responses have included revisions to terms of service, staffing increases in moderation teams, and (in Meta's case) the creation of an independent oversight board.[3] These have generally been seen as insufficient.

Despite widespread credence given to the theory of "surveillance capitalism", which posits that platforms' advertising-based business models are the root cause of such harms, regulations targeting platforms' collection and use of personal data (such as the General Data Protection Regulation) do not appear to have mitigated them.[4] The Online Safety Bill takes a different approach, by establishing duties of care for platform owners towards their users.

The apparent simplicity of this solution is deceptive: the Bill itself is lengthy and complicated. It has had several iterations, outlasting two prime ministers and six Secretaries of State for Digital, Culture, Media and Sport. Policy issues which predate the ascendancy of "big tech" – notably the circulation of child sexual abuse material (CSAM) online and the use of digital technology to facilitate acts of violent online political extremism (VOPE) – have been brought into its scope, along with provisions intended to combat fraudulent advertising and safeguard freedom of expression.

Our objective with this brief is twofold. Firstly, we aim to offer policymakers an overview of the Bill in its current state (Section 3), and a summary of the criticisms different stakeholders level against it (Section 4). Secondly, we aim to defend the duty of care principle underpinning the Bill, while making a case that the technical solutions it prescribes for mitigating CSAM and VOPE should be removed – not least because we believe other policy levers will be more effective (Section 5).

---

[1] Human Rights Council Staff. 2018. *Report of the Independent International Fact-Finding Mission on Myanmar*. Human Rights Council.

[2] BBC News, "Molly Russell: Meta and Pinterest executives must appear at inquest", 2 September 2022, available at: https://www.bbc.com/news/uk-england-london-62772754

[3] The Oversight Board's charter is available at https://oversightboard.com/governance/

[4] The theory is developed and defended in Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Profile Books. For a critique, see Gilbert, Sam. 2021. *Good Data: An Optimist's Guide to Our Digital Future*. Welbeck Publishing.

What exactly are "online harms"?

There is no settled definition. The Online Harms Whitepaper which preceded the Online Safety Bill specified 23 harms, but noted that its list was "neither exhaustive nor fixed".[5] The term is usually understood to encompass both new forms of harm arising from the diffusion of digital technology, and established forms of harm to which digital technology has added new dimensions.

Examples of new forms of harm:

- *Doxxing*: maliciously publishing personal information about a specific individual on the internet
- *Trolling*: posting unsolicited comments online with the intention of causing hurt or provoking an emotional reaction from an individual or group
- *Cyberbullying*: the use of digital messaging applications to send threatening or insulting messages to an individual
- *Cyberflashing*: sending unsolicited sexually explicit images to another person
- *Revenge porn*: posting sexually explicit images of a former partner online without their consent in order to cause them distress

Examples of established harms to which digital technology has added a new dimension:

- Child sexual exploitation and abuse
- Violent political extremism
- Incitement of violence

The Bill puts harms relating to data protection and cybersecurity out of scope, while recent comments by Secretary of State for Digital, Culture, Media and Sport, Michelle Donelan, suggest the "legal but harmful" category may be scaled back or removed. [6]

3. Overview of the Online Safety Bill

The Bill aims to:

- Make it harder for violent online political extremists (VOPE) to proselytize and recruit
- Reduce the online circulation of child sexual abuse material (CSAM)
- Mitigate risks relating to "legal but harmful" content (for example: content that might encourage eating disorders)
- Make it harder for children to access pornography
- Prevent fraudulent adverts

[5] Online Harms Whitepaper, December 2020, available at: https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper
[6] BBC News, "Online Safety Bill to return as soon as possible", 21 September 2022, available at: https://www.bbc.com/news/technology-62908598

It targets major social media platforms, search engines, and pornography websites (collectively referred to as "service providers"), which will be subject to regulation by Ofcom.[7] The regulator will set specific codes of practice for service providers, and have the power to fine them up to 10 percent of global annual turnover or £18 million (whichever is the greater) – or to block them entirely in cases of repeated violations.

The Bill's main mechanism is the creation of duties of care for service providers. Specifically, they will be required to:

- Protect their users
- Remove illegal material (including CSAM and VOPE content)
- Create and implement more robust policies on "legal but harmful" user behaviour and content (including harassment, disinformation, and self-harm material)
- Offer users more controls (for example, enhanced features to block and report other users)
- Prevent fraudulent adverts being posted
- Safeguard pluralism in debate

In addition, it requires service providers to prevent children from being able to access pornography, via age verification or other means[8]. Finally, "as a last resort" it gives Ofcom the power to mandate messaging services to scan public traffic for VOPE content, and both public and private traffic for CSAM.[9]

The Bill is framework legislation and does not specify what constitutes "legal but harmful" content, neither does it describe in detail how service providers should discharge their duties of care, or which service providers are in scope. However, Ofcom has indicated it expects to regulate 30-40 service providers in total.[10]

4. Criticisms of the Bill

Despite broad agreement that online harms are a real issue and that self-regulation by service providers is not a satisfactory solution, the Online Safety Bill has many critics. They fall into one or more of three groups: those who believe the Bill goes too far; those who believe it does not go far enough; and those who think some aspects of it are unworkable in practice.

Those who believe it goes too far are concerned that the Bill is illiberal, in that it limits freedom of expression, enables the expansion of the state's surveillance capabilities, and distances the United Kingdom from multilateral efforts at internet regulation.[11] The duty of

[7] For the avoidance of doubt, Internet Service Providers (ISPs) are not within the Bill's scope.
[8] Age verification for online services has been proposed several times before and abandoned once ministers and officials start to examine the cost, complexity and likely side-effects.
[9] Department for Digital, Culture, Media & Sport, Online Safety Bill Factsheet, 19 April 2022, available at https://www.gov.uk/government/publications/online-safety-bill-supporting-documents/online-safety-bill-factsheet
[10] Ofcom's Impact Assessment for the Online Safety Bill is available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1061265/Online_Safety_Bill_impact_assessment.pdf
[11] These include the Declaration for the Future of the Internet, the Santa Clara Principles on content moderation, the Internet and Jurisdiction Policy Network toolkits on cross-border content regulation, and the

care with respect to "legal but harmful" content is said to create stronger incentives for service providers to over-police social media posts, effectively censoring their users, and potentially producing a chilling effect on free speech online. At the same time, Ofcom's power to mandate the use of "proactive technologies" is seen by some as a back door to the introduction of a general monitoring obligation – the blanket monitoring of users' online activity by service providers, which is prohibited under the European Convention on Human Rights.

Those who believe the Bill does not go far enough are concerned that it leaves gaps – particularly in terms of the protections afforded to women and children, and the scope of the service providers to which it applies. Secondary legislation will be required to specify the "legal" harms service providers will be required to mitigate, meaning issues such as the online abuse and harassment of female public figures may not actually be addressed despite the concern and undoubted harm this causes. Meanwhile, online gaming platforms, which are used by 86 percent of 12 to 15 year olds, do not appear to count as service providers by the Bill's definition. Given that design features and aspects of gaming business models have been shown to expose children to financial harms, as well as abuse by older gamers, it is argued that they should be subject to the same duties of care as social media platforms.[12] It has been suggested that widening the scope of the Bill to encompass all digital services would enable Ofcom to tackle this issue, as well as to mitigate the risk of harms on fast-growing platforms that are currently below the Bill's scale threshold (cf infra).[13]

Those who think the Bill is unworkable point to its length, complexity, dependence on secondary legislation, and the operational challenges and costs of implementing its requirements – a process which is not expected to begin until mid-2024.[14] It is argued that – in contrast to physical injury – there is no objective way of ascertaining that emotional or psychological harm has occurred, making it impossible to determine whether service providers have discharged their duties of care.[15] At the same time, controversies of interpretation are said to be a likely consequence of relying on flexible standards and introducing categories such as "legal-but-harmful" content and "content of democratic importance".[16]

5. Commentary

We are sympathetic to many of these criticisms, but nevertheless regard the Online Safety Bill as an important policy intervention. In the following two sections, we lay out a defence of the duty of care principle on which the Bill is based, explain why we believe "proactive

---

Council of Europe guidelines on combating hate speech. See Burns, Heather, "Fixing the UK's Online Safety Bill", July 2022, available at https://webdevlaw.uk/2022/07/21/fixing-uks-online-safety-bill-part-0-not-fixable/

[12] Parentzone, "7 key issues from the Online Safety Bill report", available at https://parentzone.org.uk/article/seven-key-issues-from-the-online-safety-bill-report

[13] Woods, Lorna, Perrin, Will, & Walsh, Maeve, "The Online Safety Bill: Our initial analysis", Carnegie UK, March 2022, available at https://www.carnegieuktrust.org.uk/blog-posts/the-online-safety-bill-our-initial-analysis/

[14] Ofcom, "Online Safety Bill: Ofcom's roadmap to Regulation", July 2022, available at https://www.ofcom.org.uk/__data/assets/pdf_file/0016/240442/online-safety-roadmap.pdf, p.19

[15] Smith, Graham, "Reimagining the Online Safety Bill", August 2022, available at https://www.cyberleagle.com/2022/08/reimagining-online-safety-bill.html

[16] Written evidence submitted by Daphne Keller to the Draft Online Safety Bill Joint Committee, September 2021, available at: https://committees.parliament.uk/writtenevidence/39173/pdf/

technologies" are unlikely to be effective at tackling CSAM and VOPE, and suggest alternative policy approaches.

5.1 Service providers' duty of care

In defending the duty of care principle, we take inspiration from the political philosopher Judith Shklar, who argued that liberal societies ought to give higher priority to the avoidance of cruelty than to the promotion of social goods.[17] Many online harms involve the use of digital technology to inflict cruelty – so, by making digital technology widely available, low-cost, and easy to use, service providers have significantly amplified our capacity to inflict cruelty on one another.[18]

From a Shklarian perspective, the Bill has one great strength: it clearly gives service providers responsibility for mitigating the risks of cruelty and other harms that their business activities produce. But it is a weakness of the Bill to suggest that harm-mitigation can be achieved without trade-offs in terms of freedom of expression. For ideological, political, and financial reasons, service providers have typically been reluctant to remove content and suspend accounts unless the law has been broken or their own terms of service otherwise violated. "When in doubt", Meta CEO Mark Zuckerberg has written, "we always favor giving people the power to share more."[19] The threat of sanctions introduced by the Bill signals the opposite: that when in doubt, service providers should err on the side of caution and remove or restrict content, accounts, and features that could cause harm to other users.

We agree that this is highly likely to result in the over-policing of legal expression on social media, particularly as what counts as "harm" is not settled. There will be more instances of inconvenience and unfairness: users posting in good faith may find their access to services being shut off temporarily or permanently, or even that they have been reported to law enforcement. It is entirely predictable that features designed to facilitate reporting of abuse and blocking of abusive accounts will be used maliciously to harass opponents.

However, we believe the harms are sufficiently serious that the trade-off is worth making. In Shklar's terms, if we are presented with a choice between a system that maximizes freedom of expression and one that minimizes cruelty, we should choose the one that minimizes cruelty. Limiting freedom of expression in mass-market social media is a price we may well be prepared to pay if we are serious about mitigating online harms. Note that platforms such as Facebook/Meta and YouTube already use their terms of service to take down not just illegal child sex abuse material but animal cruelty, videos of gangland killings and much else; yet there are some forms of cruelty, such as revenge porn, over which the industry drags its feet. A focus on cruelty may help clarify boundaries. Moreover, it is important to note that such limits

[17] Shklar, Judith, "The Liberalism of Fear." In *Liberalism and the Moral Life*, ed. Nancy Rosenblum. Cambridge, Mass.: Harvard University Press. Although she died in 1992, Shklar's work has been fruitfully applied to questions of internet regulation; see for example Runciman, David, "Political Theory and Real Politics in the Age of the Internet." Journal of Political Philosophy 25(1): 3–21

[18] The argument outlined in this section is developed in Greene, Amanda and Gilbert, Sam "More Data, More Power?: Towards a Theory of Digital Legitimacy", abstract available at https://ethics.harvard.edu/event/more-data-more-power-towards-theory-digital-legitimacy

[19] Mark Zuckerberg, "Building Global Community". *Zuckerberg Transcripts* 989 (2017).

do not amount to the removal of individuals' basic right to speech. Even when permanently banned from one platform, users remain free to post on other platforms, or to build and publish on platforms of their own. It is providers of infrastructure services such as web hosting, payments, and cloud security who have the power to cut off altogether individuals' and organizations' ability to express themselves online – a power that is rarely exercised[20]. As noted, these providers are outside the Bill's scope.

There is one caveat. The duty of care to users creates high compliance costs, which only large organizations can reasonably be expected to bear. If, as some have suggested, it was extended beyond the 30-40 service providers Ofcom appears to have identified (cf supra), there would probably be adverse consequences for competition: smaller providers might become financially unsustainable, reinforcing the market power of big tech companies. For this reason, we are in favour of a scale threshold. However, we also favour bringing major gaming platforms such as Roblox into scope.

Finally, while we recognize that the Bill creates legal uncertainty, we think there are grounds for cautious optimism about the effectiveness of flexible standards in reducing harms. Since 2006, UK financial services firms have had a duty to ensure they are "treating customers fairly" (TCF). While it has often been criticized for being a vague standard rather than an enforceable rule, in practice TCF has led to firms giving more consideration to the meaning of "fair" treatment and how it can be evidenced. Together with dialogue between the regulator and firms over enforcement notices, this appears to have had a positive impact on the customer-centricity of sales practices, new product development, and organizational culture.[21] What TCF has done for financial services, an Online Safety Bill with a focus on cruelty might do for digital ones.

5.2 Child protection

Some online services, such as Gmail and Facebook, already scan communications for images that are already known to be illegal, including images of sexual abuse. US firms are required by US law to report such material to the US National Center for Missing and Exploited Children (NCMEC) and cannot report them anywhere else, which will cause a problem if Ofcom wants them reported directly to a UK agency. NCMEC currently reports about 100,000 such images to the National Crime Agency (NCA) every year resulting in several hundred prosecutions per month for indecent image offences.

Recently, some services have also started using AI to scan for new images that might be illegal. This has a much higher error rate; there are both false negatives (abuse images that escape detection) and false positives (legal images that are wrongly flagged as abusive). For example, a father who took a picture of his infant son's inflamed penis at the request of the nurse at a

[20] The most recent example of this power being exercised is described in Prince, Matthew, "Blocking KiwiFarms", Cloudflare Blog, September 2022, available at https://blog.cloudflare.com/kiwifarms-blocked/

[21] See for example Ring, John Patrick et al, "Taking notice of risk culture – the regulator's approach", Journal of Risk Research, Volume 19, 2016 and Burdon, W.M., Sorour, M.K., "Institutional Theory and Evolution of 'A Legitimate' Compliance Culture: The Case of the UK Financial Service Sector." J Bus Ethics 162, 47–80 (2020)

medical practice got a visit from the police, and lost access to his Google accounts, after the company's AI flagged the photo as abusive.[22]

The European Commission now proposes, in its Child Sex Abuse Regulation currently before the European Parliament, to extend AI-based scanning from images to text, and to insist that end-to-end encrypted services such as WhatsApp build scanners into their app, so that text can be scanned before it is sent, or after it is received. Ofcom has consulted on similar proposals, and on extending text scanning from 'grooming' to terrorist radicalisation.

Extending scanning from known-bad images to suspect images and then to text will greatly increase the number of false alarms and thus the number of innocent people caught up in the surveillance dragnet. The European Commission admitted internally that there might be a false alarm rate of 10%, but claimed that if there were 1,000,000 grooming messages, then 100,000 false alarms could be dealt with[23]. However, there are about 10,000,000,000 text messages every day in the EU and with a false alarm rate of 10%, the number of false alarms is not 100,000 but 1,000,000,000. Europe's 1.6m police officers would each have to scan 625 of them every day. Such a system would be simply unworkable.

A deeper objection is that this is the wrong type of initiative to help with the prevention and detection of violent sexual crimes against children (or violent political extremism, to which we will come in the next section). Both types of crime are embedded in local communities; detection is a task for local police forces, while prevention also involves teachers, parents, social workers and community leaders. Automated online scanning cannot substitute for local knowledge and social context.

Violent crime against children is largely family violence; worldwide, there are about 100,000 homicides of children a year. About 200 of these are in the UK; the figure rose slightly in 2020–21, possibly because of the lockdowns. The typical perpetrator is the mother's partner, although from puberty onward a growing number of homicides are committed by acquaintances. Child homicides are the visible tip of a largely invisible iceberg of child abuse, of which by far the most common kind is simple neglect. This is associated with multiple deprivation: with unstable families, poor living conditions, structural unemployment leading to endemic poverty, exacerbated by alcohol and drug abuse. It should surprise no-one that patterns for sexual violence are similar.

Abusers increasingly use mobile phones and other devices to monitor their victims. They may take indecent images of their victims or even photographs of actual abuse as a means of extortion or control. By no means all the victims are minors; a woman trying to escape an abusive husband can find him using technology to track her down and threaten her, and her children. Technical policy options must therefore be assessed in the context of abusive families. Some tech companies have started trying to redesign products to be more resistant to intimate

[22] Kashmir, Hill, "A dad took photos of his naked toddler for the doctor. Google flagged him as a criminal", *New York Times,* 21 August 2022, available at https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html

[23] Sebastian Meineck, Markus Reuter and Andre, "EU-Kommission nimmt hohe Fehlerquoten bei Chatcontroller in Kauf", Netzpolitik.org, June 2022

partner abuse, but this is difficult: the abuser can compel the victim to install spyware on their phone, share passwords and so on.

Once children reach puberty, abuse by and of peers becomes an issue. Even in the absence of abuse, about a third of teens send explicit images to each other as part of the normal process of flirting.[24] When relationships break up, one of the participants may either disseminate an image of the other without consent ('revenge porn' or 'non-consensual intimate imagery') or threaten to do so ('sextortion'). About 5-10% of children and young people report victimisation; the offenders are usually peers of the victim, with about 2% (almost all male) offending. About 10% of teens experience physical dating violence and about 15% experience some sexual victimisation.[25] Because of the volume of offending, it would be best if parents and teachers could deal with all but the serious cases; they are the responsible adults who should be shaping the behaviour and attitudes of children and young people close to them. However, in Britain, the mere possession of an indecent image of a minor has been an offence since the Protection of Children Act 1978, which makes it difficult for teachers to deal with sexualised bullying. Indeed in Scotland, all violence must be reported to the police rather than being dealt with by teachers.

Sextortion is also used by criminal gangs to target vulnerable young people to draw them into prostitution. The victims in such cases are often from dysfunctional families and may be absent from school. There is also abuse by men in authority such as priests, sports coaches, teachers and police officers. The vetting and barring regime was established after the Soham murders to mitigate these risks.

In short, child protection is a complex problem, embedded in local communities, and requires coordinated action by parents, police, social workers, schools and others.

There was an unsuccessful attempt by the Blair government to find a technological 'silver bullet' in the form of the eCAF system – a database recording all the concerns that any doctor, teacher or other public-sector worker had about any child in England or Wales, so they could be presented in a standard format to social workers. This was yet another public-sector IT disaster. On taking power in 2010, the Cameron government set up the Munro Review which recommended its abolition, as social workers were having to spend all their time sitting at screens and 'feeding the beast' rather than talking to their clients. This failing system was duly taken out of service. It would be foolish to build yet another system that burdens an already stretched public service with too many false alarms. The policy problem to fix is the inadequate resourcing of the frontline social care staff.

Another objectionable aspect is the proposal, by Ofcom and GCHQ, that end-to-end encrypted services such as WhatsApp should be required to scan private messages for illegal images and for potential grooming content on each user's phone or laptop, scanning sent messages before

[24] Charles Hymas, "Sending nude photos is 'the new flirting' for teenagers", *The Telegraph*, 23 June, 2022, available at https://www.scotsman.com/education/third-teens-have-been-sent-sexually-explicit-images-1464927

[25] Safe Lives, "Young people and domestic abuse", July 2017, available at https://www.basw.co.uk/resources/safe-young-lives-young-people-and-domestic-abuse

they are encrypted and received messages after they are decrypted.[26] This ''snoopers' charter" has already met with resistance not just from the Opposition but within the Conservative Party (amendments 152 and 153 by David Davis, Kevin Hollinrake, Philip Davies and Steve Baker to remove encrypted services from the scope of technology notices). It would be a significant departure from British law and practice (e.g. the 2016 Investigatory Powers Act's prohibition of bulk interception against UK residents). It would also fall foul of the European Court of Human Rights which maintains a staunch position against bulk surveillance without warrant or suspicion. It would also be ineffective, as the error rates for text scanning are so high that the police would be swamped in false alarms.

The only known effective way to detect grooming and new CSAM is by user reporting. The problem here is that dealing with user complaints costs money, so the large service providers make it inconvenient to contact a moderator. Here the Online Safety Bill can help by mandating that services provide an effective way for users to report illegal content and have it rapidly taken down. Tech companies already do this for copyright owners; the law should compel them to treat vulnerable users, such as women and children, with the same consideration. When such users encounter harassment or abusive material they must be able to contact moderators quickly so that they can preserve evidence, remove illegal material and block people who are trying to exploit them[27].

The scanning of private text messages has also been mooted in the context of combatting terrorist radicalisation and recruitment, so we consider that next.

5.3 Violent online political extremism

For the last ten years, the Metropolitan Police have run an Internet Referral Unit that looks for terrorist material online and asks service providers to take it down. Extending this service from material published online to private material on people's phones and laptops would be a major extension of police power. Reconciling it with human-rights law would likely be impossible.

A growing body of scholarship challenges the approach to terrorism taken by the intelligence and defence community since 9/11, as it implies there are more effective tools than extensive surveillance. One finding in this research is the very strong link between misogyny and violent political extremism, which extends across both Islamist and far-right violence. As one striking example, Joan Smith has studied all the terrorist murders in Europe since 9/11 and a significant number of mass shootings in the USA. In the great majority of cases, the killer committed a violent crime against a female acquaintance or family member before going on to murder members of the public. The U.S. scholar Val Hudson has researched the same phenomenon, and writes of the 'third day effect': on the first day of a mass shooting you hear about the killer, on the second day you hear about his victims and on the third day you hear about the girlfriend or

[26] Dan Milmo "UK could force messaging apps to adopt new technology to tackle abuse images", *The Guardian*, July 6, 2022, available at https://www.theguardian.com/technology/2022/jul/06/uk-could-force-messaging-apps-to-scan-for-child-sexual-abuse-images

[27] The EU's new Digital Services Act will give such rights to its residents, so the tech majors will have to build the machinery. It would be an embarrassment if they were not compelled to deploy it in the UK too.

female relative he beat up or killed first. In addition, both terrorists and mass shooters were very often raised in households with a hyper-dominant and abusive father. These findings imply that if there are two hundred disaffected youths attracted to a mosque with a fiery Salafist preacher, the one to watch is not the one who downloaded a copy of the Anarchist's Cookbook but the one who beat up his sister.

Other factors do matter, from a violent father through mental illness to life crises, but violence against women appears to be the strongest signal. This once more indicates that local police work, not centralised electronic surveillance, is the more effective way to prevent harm. Much the same holds for extreme right-wing violence, which since 2018 has surpassed Islamist extremism in terms of reported UK offences. The details will vary from one part of the UK to another, depending on whether there are impoverished immigrant communities, whether young men can get jobs, and (above all) whether they can start families and settle down. The details will also depend on any local history of extremism, on local police culture, and on the extent to which other guardians such as teachers, social workers, and youth clubs can support and nurture alienated young people.

It is hard to see any case for breaking everyone's privacy, in contravention of the settled British tradition, in order to intercept communications at a greater scale when more effective routes are available. And there is also a real risk that such a policy choice would feed the narratives on which extremism relies – that the government doesn't care about its citizens, that it spies on them, that it considers itself above the law, that officials lie and ministers are corrupt, and so on. MPs and ministers should consider the costs and benefits to their constituents, and to the institutions on which we rely for maintaining trust in a democratic society.

Finally, when dealing with terrorism, it is counterproductive to talk it up and use it as an excuse. In the words of Salman Rushdie, "How to defeat terrorism? Don't be terrorised. Don't let fear rule your life. Even if you are scared."

6. Recommendations

We have argued that the Online Safety Bill is right to impose a duty of care on important digital platforms. We have also challenged some of its current provisions as having costs that outweigh their purported benefits, especially when there are more effective – offline – ways to tackle the harms. We end here with some specific recommendations:

1. The Bill's scope should be extended to gaming service providers. Gaming platforms expose children to the same risks of abuse as social media, as well as to financial harms.
2. The power for Ofcom to mandate the use of "proactive technologies" should be dropped from the Bill. Client-side scanning is technically ineffective and impractical as a means of mitigating violent online political extremism and the circulation of child sexual abuse material, and undermines fundamental freedoms. In general, Ofcom must not be able to

mandate the use of 'accredited technologies' but rather regulate for the desired outcomes.

3. There is no technological "silver bullet" that can solve violent political extremism and child sexual abuse. These harms are better tackled at local community level through increased funding of policing and social work. Recruitment of more police is already in hand, but there needs to be more focus on family violence, and better recording of men who threaten, or commit violent crimes against, women and children – not only as a child-protection measure but as an early warning for other types of violent crime, including terrorism.
4. However, the Bill should require all firms offering regulated user-to-user services to enable users to contact them easily in order to block other users who are harassing or attempting to exploit them; to obtain the rapid removal of illegal material; and to secure evidence.
5. Teachers should be given a liability shield by the Crown Prosecution Service with respect to indecent images, so they are better able to help deal with the less serious cases of online harms like sextortion.

7. About the authors

Ross Anderson is Professor of Security Engineering at the Universities of Cambridge and Edinburgh. He is a Fellow of both the Royal Society and the Royal Academy of Engineering, and chairs the Foundation for Information Policy Research.

Diane Coyle is Bennett Professor of Public Policy and Co-Director at the Bennett Institute for Public Policy, University of Cambridge. She is also a Director of the Productivity Institute, a Fellow of the Office for National Statistics, an expert adviser to the National Infrastructure Commission, and Senior Independent Member of the ESRC Council.

Sam Gilbert is an affiliated researcher at the Bennett Institute for Public Policy and the author of *Good Data: An Optimist's Guide to Our Digital Future*. He was previously Employee No.1 and Chief Marketing Officer at the fintech unicorn ManyPets, and held senior roles at Experian and Santander.

UNIVERSITY OF CAMBRIDGE

www.bennettinstitute.cam.ac.uk

Bennett Institute for Public Policy Cambridge