Safety without Censorship

A better way to tackle online harms

By Caroline Elsom

Centre for Policy Studies

About the Centre for Policy Studies

The Centre for Policy Studies was recently named by Conservative MPs polled by ComRes as the most influential think tank in Westminster. Its mission is to develop policies that widen enterprise, ownership and opportunity, with a particular focus on its core priorities of housing, tax, business and welfare.

As an independent non-profit think tank, the CPS seeks likeminded individuals and companies to support its work, but retains editorial control of all of its output to ensure that it is rigorous, accurate and unbiased.

Founded in 1974 by Sir Keith Joseph and Margaret Thatcher, the CPS has a world-class track record in turning ideas into practical policy. As well as developing the bulk of the Thatcher reform agenda, it has been responsible for proposing the raising of the personal allowance, the Enterprise Allowance, the ISA, transferable pensions, synthetic phonics, free ports and many other successful policy innovations.

About the author

Caroline Elsom is a Senior Researcher at the Centre for Policy Studies, who specialises in technology, education, business and tax reform. She has written for CapX about online harms and given evidence before the House of Lords Select Committee for Democracy and Digital Technology. Caroline was previously a Senior Parliamentary Researcher to a Home Office minister with responsibility for online safety. She has also been a Press Officer at Conservative Campaign Headquarters.

Acknowledgements

I would like to thank the various ministers and MPs, special advisers, business leaders and policy experts we spoke to in the preparation of this work. Particular thanks are due to the Coalition for a Digital Economy (Coadec) and 89up who supported the report. I would like to thank Lucky Dube for his research assistance, and Robert Colvile and Alex Morton for their editing guidance.

Contents

Executive Summary

- The Online Harms White Paper as currently being discussed is both unworkable and risks damaging free speech and competition online.

- We propose a new model of regulation where illegal and legal content is clearly delineated with Ofcom serving as regulator.

- The model includes beefed up resourcing for police to enforce laws tackling egregious harms like Child Sexual Abuse & Terrorist content.

- It would also ensure that any decisions on restricting free speech online was in the hands of elected officials not unelected regulators - whilst ensuring that a clear system was in place to address emerging challenges online.

In spring 2019, the Government published the Online Harms White Paper, setting out its plan to regulate online content.

It proposed a new statutory duty of care to make companies take more responsibility for the safety of their users and tackle harm caused by content on their services. Compliance with this duty of care would be overseen and enforced by Ofcom, who would have a suite of powers to take enforcement action against companies whose processes were not effective in combatting harm.

The core aims were to protect the most vulnerable in society (namely children), protect national security, protect freedom of expression and to promote business in the UK. A vision of a safer, kinder and more prosperous internet.

However, the current policy trajectory for online harms regulation will cause confusion for the online technology ecosystem that will seriously threaten freedom, privacy, competitiveness and the UK's reputation for democratic accountability.

The vagueness of the responsibilities under the duty of care, of defining 'harm' and the problem of using the same overarching regime for both illegal and legal acts clouds the effectiveness of handling the issues with either. What makes this proposed sacrifice all the more troubling is that the regulation in the White Paper is unlikely to make us any safer.

We found the following:

- Regulation of speech that is perceived as harmful, but remains legal, raises serious free speech concerns and increases the risk of abuse of power in the future.

- Rules implemented by private bodies can be arbitrary and may be difficult for users to hold to account.

- Compelling companies to take down harmful content is overemphasised in the proposals without enough due care given to changing processes that lead to these harmful environments existing.

- There is a lack of evidence to support some claims about the impact of legal 'harmful' content and cancelling content does not address the research-deficiency that exists.

- Codes of practice with prescriptive solutions are easier to measure but limit the ability to adapt when new problems arise and do not address the root causes of problems.

- The imposition of a duty of care and a broad regulatory framework on tech companies of all sizes will inhibit innovation and competition in one of the UK's most important sectors of growth.

This paper lays the groundwork for a more targeted and effective model of regulation, in which free speech whether it is offline or online is treated fairly and it is Parliament, rather than regulators, who set the parameters for what is legal and/or harmful and what is not.

We propose:

- A tough new regulator, still under the oversight of Ofcom, but one that should work collaboratively with the police and the Crown Prosecution Service to tackle the scourge of criminal activity online.

- A clearly demarcated regulatory regime for legal vs illegal content. The dividing line offers much greater safeguards against overreach.

- Significant new resources for the police to conduct forensic investigations online and ensure cases can be brought by prosecutors that ensure egregious illegal acts will be punished properly.

- In addition, harms that occur online that are still lawful will be identified and reported to the regulator by stakeholders with relevant expertise and bodies designated to lodge 'Super Complaints' with the regulator.

- The regulator should provide regular thematic reports and thought leadership to Parliament who will in turn make recommendations to the Government on additional legislation to address these challenges. This will serve as a vital tool for connecting online harms to democratic scrutiny.

> **“We have entered an age where anyone with a smartphone walks around with their own printing press, broadcast station and meeting hall in their pocket.”**

We have entered an age where anyone with a smartphone walks around with their own printing press, broadcast station and meeting hall in their pocket. Being online has become an essential tool for the exercise of democracy and for accessing goods and services. Building a regulatory framework to manage this complex world requires trade-offs – it is a question of priorities and practicalities. Our model does not solve every problem or pitfall, but nor, we must accept, can any. We propose a more effective and democratic model that is fit for the growing digital economy and that is built to last.

Introduction

Tackling online harms has been an ambition for successive leaders grappling with a fast-changing digital world.

From Instagram accounts influencing children to commit suicide to fake news on Twitter influencing behaviour during the pandemic, the political pressure to act has never been higher.

Yet despite the clamour to act, the reality of policing harms online is not about a simple separation of good content from the bad. What is now deemed to be potentially harmful online can cover anything from the most extreme illegal acts to free expression of views that some may take as offensive. A tough, fair and accountable model that incorporates this breadth is impossible to achieve.

In the UK, policymakers' answer to the perils of the internet has come in the form of the Online Harms White Paper (hereafter, the White Paper), released in April 2019. Its core aims are to make the UK 'the safest place in the world to go online, and the best place to start and grow a digital business'.[1] The plans made the UK the first major country to set out a route map for regulating content and the conduct of online business. Aims to legislate on this topic have since been set out in broad terms in the 2019 Conservative General Election manifesto, which states:

> 'We will legislate to make the UK the safest place in the world to be online – protecting children from online abuse and harms, protecting the most vulnerable from accessing harmful content, and ensuring there is no safe space for terrorists to hide online – but at the same time defending freedom of expression.'[2]

In the Queen's Speech, the Government's commitment to 'develop legislation to improve internet safety for all' was once again laid out.[3] In her last act as Secretary of State for Digital, Culture, Media and Sport, Nicky Morgan, alongside the Home Secretary, Priti Patel, released an initial response to the White Paper setting out an intention to appoint Ofcom as the appropriate regulator. There was clear direction from the new administration that it would seek to make progress in this area in the current parliament.

As with most areas of daily life, the COVID-19 pandemic forced these plans on to the backburner. It also drove an unprecedented shift in demand for online services. People of all ages, some for the first time, embraced online communications for their virtual classrooms, community organisation, business operations

[1] HM Government, "Online Harms White Paper". Available from: https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper.

[2] The Conservative and Unionist Party, "Manifesto 2019", p20. Available from: https://assets-global.website-files.com/5da42e2cae7ebd3f8bde353c/5dda924905da587992a064ba_Conservative%202019%20Manifesto.pdf.

[3] Prime Minister's Office, 10 Downing Street, "Queen's Speech December 2019". Available from: https://www.gov.uk/government/speeches/queens-speech-december-2019.

and family catch-ups. Issues around misinformation, scamming and child safety were, accordingly, pushed up the agenda.

But this involuntary interruption to business as usual also gives us an opportunity to pause and reflect. What the UK does next will have an impact across the world at a moment when both the EU's proposed Digital Services Act and reform of Section 230 in the US are in the pipeline in the near future. So this paper attempts to test whether the plans on the table will really meet the stated aims, namely protecting children and vulnerable people, protecting national security, protecting freedom of expression and promoting business.

The White Paper sets out a sweeping 'duty of care' which would give companies statutory responsibility for potential harm that legal user-generated content might cause. A new or existing regulator overseeing everything from search engines to messaging apps and even product review sections would be handed the power to impose fines and individual liability on firms' leadership teams, or otherwise block the activities of businesses it deemed non-compliant.

The big problem is that while overseas-based technology giants may be the target of these measures, the plans are likely to create a considerable hurdle for new entrants looking to challenge them. Start-ups would be forced to devote disproportionate time and resources to monitoring and filtering activity in order to avoid falling foul of rules designed around today's dominant players. A culture of over-cautiousness is likely to result, stifling freedom of expression online.

If the measures taken force firms to withdraw services users depend upon for the effective exercise of their freedom of expression, or which media companies depend on for the protection of sources and freedom of the press, then the new measures will not only have inflicted harm, but failed under their own terms too.

“If the measures taken force firms to withdraw services users depend upon for the effective exercise of their freedom of expression, or which media companies depend on for the protection of sources and freedom of the press, then the new measures will not only have inflicted harm, but failed under their own terms too.”

Such critiques of the White Paper do not seek to underestimate the immensely difficult situation that government has faced for many years in keeping up with emerging trends online and responding to public demands for action. It is encouraging to see a genuine attempt to grapple with the multitude of conflicting rights and principles that govern individual and company behaviour online. But getting the balance right between protecting people from harm and upholding the right to freedom of expression and privacy is a delicate line to tread.

This paper argues that the White Paper's plans will not work as intended, and sets out an alternative framework of principles from which any measures to address online harms should be approached. It explores the relationship between harm, duty and legality to weigh up security and freedom on the internet. This involves looking at the current landscape and the attempts already made by various online operators to adapt to the demand for safer services.

Our regulatory model is tough on the mechanisms of harms and places democratic decision-making at its heart. At the same time, it addresses the variety and complexity of digital businesses today and deals more fairly with the online and offline worlds. This takes into account the current regulatory landscape and lessons that can be learned from recent developments abroad.

Part 1: The current regulatory landscape

In the discourse around online harms, politicians and campaign groups often compare the internet to the 'Wild West'.

A picture is evoked of an uncharted, unorganised territory where people act lawlessly and unjustly – anything goes.

However, in the nearly 30 years since the new frontier of the global internet began to be colonised, many rules, regulations and remedies have developed to govern the online world. This is also where many of the problems have arisen, as governments have struggled to keep up with the changing nature of the internet, leaving a fragmented but at the same time overlapping regulatory landscape.

There is no succinct area of law that online harms occupy, which would make the role of the proposed statutory regulator quite unlike similar fields, such as data protection and broadcasting. Even among the dominant online players, there are vast differences in what type of content is shared and how it is spread. So drawing broad responsibility for online harms under a single new regulator would create a sprawling remit with unprecedented power to wield over internet freedom and wider civic life in the UK.

A key oversight of the White Paper is that while it acknowledges that there are several regulators operating in the online harms space, it does not include detailed analysis of the dynamic effects of its new measures on these existing bodies. There are already at least five regulators crossing significantly into the sphere of online activity:

- Ofcom
- the Advertising Standards Authority (ASA)
- the Information Commissioner's Office (ICO)
- the Electoral Commission
- the Competition and Markets Authority (CMA)

Each has specific responsibilities around the use of content, data, competition and conduct that would inevitably overlap or conflict with the role of a new online harms regulator. They are also already working together in many cross-regulator initiatives dealing with different aspects of the online harms field.

> **"Online companies themselves are not simply standing still while they wait for the Government to set out its grand rulebook."**

Online companies themselves are not simply standing still while they wait for the Government to set out its grand rulebook. They are well aware of the damage to their reputation that occurs when serious harm occurs on their platforms. More than ever before, news stories are pointing the finger at the platforms themselves when users are abusing each other via their channels. It is in the companies' own interests to take steps to clean up their act to retain customers and advertising revenue.

In other words, far from being a lawless space that requires taming, the online space is already highly regulated by a plethora of different organisations. Even before self-regulation and industry-based initiatives are taken into account, there is already a complex web of controls over the actions of online businesses. This section sets out the backdrop of regulation upon which any new legislation on online harms will sit and the tensions that this may cause.

Ofcom

The Government's initial response to the White Paper confirmed that it is minded to appoint Ofcom as the statutory regulator for online harms. This follows the recommendation of the House of Commons Science and Technology Committee that Ofcom be ready and empowered to tackle online harms by the end of October 2019.[4] Lord McNally's Online Harms Reduction Regulator Bill also proposes to give Ofcom immediate powers over an even wider duty of care than the White Paper.[5]

It appears almost inevitable now that Ofcom will become the designated online harms regulator. Nevertheless, there is still ample opportunity for the Government to carefully consider exactly how far this role should go.

Ofcom is already a 'super-regulator'. It carries out the work previously done by five separate bodies, as well as exercising some functions previously reserved to secretaries of state. Ofcom currently grants licenses to online platforms that broadcast live TV and online-only linear TV. It sets the broadcasting rules on due accuracy and due impartiality. From 2010 it introduced rules for what is now more than 300 on-demand services, with a remit to protect children, prevent incitement to hatred and limit product placement and sponsorship within programming.[6] Ofcom also has jurisdiction over telecommunications firms including internet service providers, fixed line telecoms, mobile services and the airwaves over which any wireless devices operate.

> **"Giving Ofcom jurisdiction over online companies would hand unprecedented powers to a single body responsible for adjudication over free expression in both broadcast media and social media."**

Giving Ofcom jurisdiction over online companies would hand unprecedented powers to a single body responsible for adjudication over free expression in both broadcast media and social media. There are already doubts as to whether Ofcom is institutionally suitable for the functions it would be taking on.[7] Ofcom is the arbiter of harmful and offensive content on television under the Broadcasting Code, allowing it to directly censor the output of the media. If Ofcom is to take the reins of online harm regulation, it would need to establish a completely separate structure to that of its other functions to prevent overreach.

It is acknowledged in the White Paper that Ofcom could not be expected to handle the sheer volume of complaints that it expects a new regulator to be inundated with. Ofcom would instead be taking on the primary role of overseeing the requirement on relevant companies to have the appropriate terms and conditions

[4] House of Commons Science and Technology Committee, "Impact of social media and screen-use on young people's health", p69. Available from: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf.

[5] Lord McNally, "Online Harms Reduction Regulator (Report) Bill". Available from: https://publications.parliament.uk/pa/bills/lbill/58-01/022/5801022_en_2.html#l1g1.

[6] Ofcom, "Addressing harmful online content", p15. Available from: https://www.ofcom.org.uk/__data/assets/pdf_file/0022/120991/Addressing-harmful-online-content.pdf.

[7] Stefan Theil et al., Bonavero Institute of Human Rights and the University of Oxford Law Faculty, "Response to the public consultation on the Online Harms White Paper", p9. Available from: https://www.law.ox.ac.uk/sites/files/oxlaw/bonavero_response_online_harms_white_paper_-_3-2019.pdf.

accompanied by an effective internal complaints processes.

While the initial response commented on feedback about 'super-complaints' being made directly to the regulator, it gave no indication about whether this idea would be taken forward and in what form. By their nature, complaints about online harms are likely to be far more complex and context-dependent than other areas Ofcom already deals with, such as broadcasting, where communication is largely a one-way process. Complaints handling is therefore another area where Ofcom would need to take extra care to make sure its new functions would not be shaped by existing structures.

> **"By their nature, complaints about online harms are likely to be far more complex and context-dependent than other areas Ofcom already deals with, such as broadcasting, where communication is largely a one-way process."**

Earlier this year the Government confirmed the appointment of Ofcom as the regulator for video sharing platforms under revisions to the EU Audiovisual Media Services (AVMS) Directive. Enhanced powers which apply to video on-demand services are being extended over video sharing platform services, including those who do not have 'editorial responsibility' for the content they host. Sites whose 'principle purpose' is video sharing (YouTube, Vimeo, etc) are in scope, but others could fall in scope if a 'dissociable section' or 'essential functionality' of their platform is devoted to such content.

Specifically, video sharing platform services will be required to put in place restrictive measures to protect minors from harmful content and to protect the general public from incitement to violence or hatred and content constituting criminal offences. Although the continued implementations of the AVMS Directive would be assessed at some point following the end of the transition period now that the UK has left the EU, Ofcom's powers in this area will apply from the autumn. Ofcom is therefore likely to have boosted powers over video services in advance of any new and duplicating legislation on online harms.

Ofcom's latest initiative in preparation for taking on a new remit for online harms has been to set up the Digital Regulation Cooperation Forum in conjunction with the ICO and the CMA. While dialogue between these organisations is of course positive, it is unclear what this new non-statutory body will add that is not already occurring in the normal course of their work, especially when there is a mushrooming number of similar coordinating initiatives, just a handful of which are mentioned below.

The Advertising Standards Authority

The ASA is the independent regulator of advertising across all media which applies the Advertising Code set out by the Committees of Advertising Practice. Among its aims is to hold technology companies to account for improving their standards of brand safety and advertising fraud, making sure that digital advertising does not unwittingly support harmful content. For example, there have been high-profile instances of brands running ads on video sharing platforms that host terrorist content or images of child sexual abuse. In June it launched a new Scam Ad Alert system in partnership with digital advertising and social media platforms.[8]

[8] Advertising Standards Authority, "We've launched a Scam Ad Alert system to help better protect consumers online". Available from: https://www.asa.org.uk/news/ASA-launches-scam-ad-alert-system-to-help-better-protect-consumers-online.html.

The Incorporated Society of British Advertisers (ISBA) has already flagged concerns that the White Paper may adversely impact the existing cooperative advertising regulatory system by introducing overlapping measures with no clarity over who would have regulatory superiority.[9] The ASA has been developing its own measures to keep up with changes to digital advertising through its 'More Impact Online' five-year strategy that focuses on harnessing artificial intelligence (AI) and machine learning against online harm. There is further concern that new regulation would impact on the existing regulatory framework for age-restricted advertisements.[10]

The Information Commissioner's Office

The Information Commissioner's Office has an ever-expanding role that crosses into the online harms space. Data regulation is crucial to many core issues, including the use of personal data through profiling, cross-device tracking (which allows business to track website users across all their devices) and then the targeting that drives platforms' delivery of online content.

It also has responsibility over data protection and compliance with the EU's General Data Protection Regulation (GDPR). Despite the many issues with the GDPR, its introduction has made businesses more conscious of data protection, the need to explain algorithmic decision-making more clearly and the public's right to know information about themselves. The Centre of Information Policy Leadership noted that the GDPR has made wide-ranging improvements to organisations' data accountability and transparency arrangements, increasing trust in how organisations handle data in the digital age.[11]

A new model of accountability for online harms will likely encounter similar problems of scalability, proportionality and flexibility to those that the ICO has faced with the GDPR. Lessons urgently need to be learnt about the ability of large technology firms to adapt to the GDPR versus that of smaller companies. New rules must avoid adding another separate regulatory regime that adds to the already heavy burden of compliance. Companies from online games producers, to mobile marketers, to new social networks took flight from Europe over the GDPR.[12] A similar effect with online harms regulation should be avoided at all costs if those smaller players already creaking over the weight of the GDPR compliance are expected to weather another storm in the post COVID-19 economic climate.

> **“A new model of accountability for online harms will likely encounter similar problems of scalability, proportionality and flexibility to those that the ICO has faced with the GDPR.”**

The GDPR experience has also flagged up a problem that will be shared with online harms regulation as to what constitutes 'private communication'. Article 2.2c of the GDPR exempts from regulation the processing of personal data 'by a natural

[9] Incorporated Society of British Advertisers, “Response to UK Government Consultation: Online Harms White Paper”, p4. Available from: https://www.isba.org.uk/media/2129/online-harms-white-paper-isba.pdf.

[10] Ibid.

[11] Centre for Information Policy Leadership, “GDPR One Year in: Practitioners Take Stock of the Benefits and Challenges”. Available from: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_report_on_gdpr_one_year_in_-_practitioners_take_stock_of_the_benefits_and_challenges.pdf.

[12] Ivana Kottasová, CNN, “These companies are getting killed over GDPR”. Available from: https://money.cnn.com/2018/05/11/technology/gdpr-tech-companies-losers/index.html.

person in the course of a purely personal or household activity'. The same difficulty exists in dealing differently with private messaging, public broadcasting and the myriad of communications forms in-between. There will inevitably be a tension between the ICO's clear purpose of data protection and the prevention of harm, which involves some element of tracking and surveillance beyond the limits of the ICO's role.

"Smaller companies in particular are less likely to have the resources to design multiple versions of the same product in order to satisfy the rules."

Even though the UK has left the EU there is little room for it change the requirements on firms that the GDPR brought in. In order to continue the free flow of data between the UK and the EEA, it needs to secure a data adequacy agreement, which requires sign-off by the European Commission. A number of contentious areas remain, including the handling of data under the Investigatory Powers Act 2016, parts of the Data Protection Act 2018, alleged British abuse of the Schengen Information System and the recent Schrems II judgement on the EU-US Privacy Shield.[13] While these issues are ironed-out, there is a limit to what Britain can do to strike out on its own when it comes to data regulation.

A further area of regulation that has arisen from the GDPR is the ICO's Age Appropriate Design Code (AADC). It sets out 15 standards that online services should meet to protect children's data, including default 'high' privacy settings and restrictions on nudge techniques. Following its publication in January 2020, the Code has come into force as of September 2 – and after its year-long implementation period the ICO will begin review and enforcement action on companies that do not comply with the code.[14]

The AADC is flawed in many ways. It effectively mandates that every company operating online that could conceivably be accessed by someone under 18 years old must age-gate its users or sanitise all of its services. Smaller companies in particular are less likely to have the resources to design multiple versions of the same product in order to satisfy the rules. Although, as with the White Paper, it leaves discretion on the degree of enforcement with the regulator, it creates a similar problem of uncertainty for businesses on how the rules may be applied to them.

Regardless of the merits of the AADC, it is already in force. With that in mind, if age appropriateness is also brought under online harms regulation too, this would be another duplication of regimes under two separate regulators – reflecting a confusing patchwork of potential regulation.

The Electoral Commission

For the Electoral Commission, the White Paper's proposals would mean that regulation of campaign spending by political parties and other campaigners would extend to online providers of the tools they use for advertising and communicating with the public. In effect, it would create two regulatory regimes for the same issue – one for parties or

[13] Institute for Government, "UK–EU future relationship: data adequacy". Available from: https://www.instituteforgovernment.org.uk/explainers/future-relationship-data-adequacy.

[14] Information Commissioner's Office, "ICO publishes Code of Practice to protect children's privacy online". Available from: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/01/ico-publishes-code-of-practice-to-protect-children-s-privacy-online/.

other campaigners and another for social media companies. There is understandable concern from the Electoral Commission that if new laws are not designed carefully, it could serve to undermine the 'rules of the game' that form the basis of our democratic processes.[15]

For example, if the details of new regulation includes mandatory ad libraries and reports by social media platforms running political adverts, then complaints to the online harms regulator would also affect the Electoral Commission's responsibilities over transparency for members of the public and regulation of political finances. The recent Intelligence and Security Committee Report on Russia further suggested that the Electoral Commission should be given more power to "stop someone acting illegally in a campaign if they live outside the UK". As with other regulatory bodies already operating in the online harms space, it is unclear who would have ultimate supremacy in dealing with these sorts of cases. In seeking to cover the gaps in digital campaigning, more loopholes or duplicate regulation of the same issues might result.

The Competition and Markets Authority

The Competition and Markets Authority looks at the business models of online companies that may be used to discriminate against competitors or to the detriment of consumers and innovation. While not directly concerned with harmful content itself, where online companies are abusing a dominant position or misleading customers, their harmful behaviour may be investigated. The CMA has already been tackling various legal types of harm, including online gambling bonus promotions, fake online reviews, secondary ticketing and misleading online hotel booking search results.

> **"The debate over online harms too often strays into discussion about the broader behaviour and market position of today's tech giants."**

In March, it created a new Digital Markets Unit as part of its increasingly interventionist stance to tackle perceived consumer harms both online and offline. It is seeking wide-ranging powers including the ability to order Google to share click and query data (logs of user interaction with search engine results) with rival search engines, force Facebook to offer choice over whether to accept targeted advertising and to impose a 'separation of platforms where necessary'.[17]

The debate over online harms too often strays into discussion about the broader behaviour and market position of today's tech giants. The temptation to use the guise of online harms regulation to tackle competition issues should be resisted, not least because the CMA is already scoping out its own expansive powers to specifically target this issue.

[15] The Electoral Commission, "Response: Online Harms White Paper". Available from: https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/changing-electoral-law/transparent-digital-campaigning/response-online-harms-white-paper.

[16] Intelligence and Security Committee of Parliament, "Russia", p36. Available from: http://isc.independent.gov.uk/committee-reports.

[17] Ryan Browne, CNBC, "UK competition watchdog seeks to curb Google and Facebook's dominance of online advertising". Available from: https://www.cnbc.com/2020/07/01/uk-cma-seeks-to-curb-google-and-facebook-online-advertising-dominance.html.

Part 2: Understanding harm

Harm is an amorphous concept – it changes according to the views of whoever is empowered to evaluate it, or whether an individual is causing or receiving the harm.

Agreeing a single all-encompassing definition, or proscribing an exhaustive list of individual harmful behaviours or content, has long been the first hurdle that plans to legislate in this area have fallen at. Building the foundations on which online harms legislation should be built therefore requires us to break down the 'where, who and what' of harm.

Where?

First, the 'where'. What makes this area of law so difficult is that for the most part, it is not online where the worst harm is committed. The possession or communication of content leads to harm that occurs offline. With child sexual abuse, for example, it is the child being abused that is of greatest concern. With terrorist content, it is the possible attack that matters. With disinformation, it is the influence of individuals' actions in the real world. Likewise pro-anorexia forums, or content glamorising substance abuse.

In essence, the issue for policymakers is that they are frequently seeking to create a separate set of rules for the online world in order to police offline behaviours. An ever-increasing number of people claim that online content has caused them emotional or psychological harm but that could also be applied to unpleasant speech experienced offline. That is not sufficient reasoning to censor it if it is not illegal.

Who?

Next, there is the 'who'. While online harms regulation is overwhelmingly framed as being about the regulation of platforms – especially large, American-owned platforms – the target is really the activities of its users. Amid all the rhetoric about evil tech giants neglecting their moral duty to keep us safe, it is easy to lose sight of the true perpetrators of harm – the humans behind the keyboards. Individuals generate content (or design bot systems to generate it on their behalf), most of it legal free expression, a tiny proportion of which could cause harm (and most of which is already illegal).

> **"Amid all the rhetoric about evil tech giants neglecting their moral duty to keep us safe, it is easy to lose sight of the true perpetrators of harm – the humans behind the keyboards."**

The primary consequence of tougher sanctions on platforms is that they will take a much more zealous approach to removing user-generated content that could be interpreted as harmful, even if that possibility is remote. For any regulation to be effective, it must aim squarely at the root cause of harm – the minority of individual human users using the internet for nefarious purposes – without impinging upon legitimate users' enjoyment of these platforms.

What?

Finally, and most critically, the 'what'. The White Paper aims to impose a new statutory duty of care that sits above separate Codes of Practice relating to different types of harm. But breaking down exactly what counts under this collective notion is by no means clear-cut. In a recent study of eight official UK reports about platform regulation published between 2018 and 2020, close to 100 different online harms were mentioned.[18]

Which types of content would count as harmful is not defined in the White Paper – instead an 'initial' list of content and behaviours are provided. A list of at least 23 separate harms it expects to be in scope is laid out: some have a clear definition, while others are less clearly defined or only judged harmful when children are exposed to them.[19] The harms range all the way from child sexual exploitation and abuse, through to trolling and 'excessive screen time'.

Significant pressure has been put on government from certain quarters to expand this list further, in particular to include economic harms.[20] Lord McNally's Private Members Bill on online harms regulation, for example, lists fraud and financial crime.[21]

The main development in what would count as harmful came in the initial consultation response, which signalled a shift to a differentiated duty of care. For legal but harmful content seen by adults, it appears as though platforms will be allowed to set their own content standards in their terms and conditions. The regulator would then step in to ensure that these T&Cs were enforced transparently, consistently and effectively, although the wording differs slightly throughout the response.

"In a recent study of eight official UK reports about platform regulation published between 2018 and 2020, close to 100 different online harms were mentioned."

While greater differentiation between legal and illegal content is welcome, the seeming clarification that the regulator will be focusing and adjudicating on internal systems rather than specific pieces of content is (rather ironically) not as transparent as it seems. Judging the effectiveness of platforms' systems still requires the regulator to decide what standards of reducing harm the platforms are being measured against. This leads back to the definitional problem of the regulator needing to set out what counts as effective action for platforms to take against an expansive range of content types. A differentiated approach that still puts legal and illegal harms under the same duty of care is not enough.

Before setting out our alternative approach, we should go back to the core objectives that policymakers are seeking to achieve in creating a new law on online harms. Taken from the stated aims set forth in the Government's manifesto, these are: to protect children and vulnerable people, to protect national security, to protect freedom of expression and to promote business.

[18] Philip Schlesinger, London School of Economics, "The Changing Shape of Platform Regulation". Available from: https://blogs.lse.ac.uk/medialse/2020/02/18/the-changing-shape-of-platform-regulation/.

[19] HM Government, "Online Harms White Paper", p31. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

[20] For examples, see British Phonographic Industry, "BPI responds to Government's Online Harms White Paper". Available from: https://www.bpi.co.uk/news-analysis/bpi-responds-to-governments-online-harms-white-paper/.

[21] Lord McNally, "Online Harms Reduction Regulator (Report) Bill". Available from: https://publications.parliament.uk/pa/bills/lbill/58-01/022/5801022_en_2.html#l1g1.

I: Protecting children and vulnerable people

Protecting the physical safety of children is consistently the top priority of legislation on online harms. Child sexual exploitation and abuse is placed front and centre, but with a wide array of related issues particularly affecting children listed alongside, including cyberbullying, advocacy of self-harm, pornography and excessive screen time. So let us look at how regulation currently operates in this area for different legal and illegal acts under this umbrella.

Child sexual abuse

Child sexual abuse imagery (CSAI) is already emphatically illegal – it is an offence to possess, distribute, show or make indecent images of children, with the latter three carrying sentences of up to 10 years. There are a wide range of initiatives and collaborative efforts on this critical challenge already, including work through the Five Eyes countries and international self-regulatory efforts including the Technology Coalition and the Internet Watch Foundation (IWF). The IWF have been effective by building an approach to removing illegal CSAI that works on a global scale through INHOPE (the International Association of Internet Hotlines) which should be operating reporting systems in 50 countries by the end of the year.[22]

In the UK, the IWF are a recognised 'relevant authority' that can issue Notice and Takedown in partnership with the police to make sure that evidence is preserved for their investigations.[23] The White Paper hints that the new regulator would have 'oversight of the take-down of illegal content', which signals that it may seek to change the grounds for takedown or take over that power completely.[24]

> **“The current efforts work precisely because they are not perceived as a tool of any particular state or ideology.”**

The current effectiveness of these international schemes in tackling CSAI online in part comes from their limited parameters, only acting on content which is always illegal. Allowing work on this topic to be subsumed into a UK statutory regulator as opposed to being led by law enforcement and international expert bodies would give credence to repressive governments seeking use the concept of child sexual imagery to crack down on other issues they choose, such as sexuality. The current efforts work precisely because they are not perceived as a tool of any particular state or ideology. Under a statutory UK online harms regulator, which had a multitude of other harms in its remit, the clarity of purpose that underpins the work of the existing authorities, the IWF and other initiatives that address this harm already across borders might risk being undermined.

It is also worth noting given the renewed focus on social media in the Government's initial consultation response that fewer than 1% of sites sharing CSAI are social networking sites or video channels, with

[22] Internet Watch Foundation, “Online Harms White Paper Response”, p3. Available from: https://www.iwf.org.uk/sites/default/files/inline-files/IWF%20Online%20Harms%20White%20Paper%20Response.pdf.

[23] Internet Watch Foundation, “The laws and assessment levels”. Available from: https://www.iwf.org.uk/what-we-do/how-we-assess-and-remove-content/laws-and-assessment-levels

[24] HM Government, “Online Harms White Paper”, p63. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

image hosts making up 82% of the sites.[25] Efforts to tackle this particular issue through a model designed around public platform content is therefore likely to be largely fruitless. Although sometimes linked to the sharing of CSAI, the separate offence of 'sexual communication with a child' does often start where children are most accessible – their social media accounts. Crucially, this offence applies to both online and offline communication, regardless of whether it has happened via email, text message, written note or spoken word.

> **"Police forces are already overwhelmed by the content being brought to their attention by online platforms relating to the abuse of children."**

The NSPCC identified that 70% of offences for sexual communication with a child in England and Wales took place on Facebook, WhatsApp, Snapchat or Instagram in 2017-8.[26] More recently there have been growing reports of groomers using online games like Fortnite, Minecraft and Roblox to access victims.[27] Concerns have heightened during the COVID-19 restrictions that children are spending much more time online, often without supervision and often while feeling isolated and lonely.[28]

For social media platforms, the key issue here centres on the ability to accurately identify groomers and children through the private data they provide to the platforms. Data regulation falls under the purview of the ICO. Indeed, the ICO has ongoing activities that cross into the online harms space, such as the investigation into TikTok's operations looking at the use of children's private data and its open messaging system that forced it to adopt a new 'family safety mode'.[29] The industry is constantly adapting to pressure from the public and existing regulators by adding new child safety features to its platforms to keep up with emerging trends in perpetrating abuse.

Adding another layer of regulation here misses the point at which meaningful improvements could be made to child safety online. Police forces are already overwhelmed by the content being brought to their attention by online platforms relating to the abuse of children.

The attrition rate from industry referral to recorded crime to justice outcomes is sizeable, with over 70,000 industry referrals made in 2016-7 but under 3,500 convictions relating to indecent images of children. Although around half of industry referrals are later assessed by the NCA as informational requests (viral images deemed adult pornography, images of clothed children, or technical errors), the National Police Chiefs Councils lists other factors that are time- and resource-related, including the suspect having moved address, no evidence being found in a search or the information being too historic

[25] Internet Watch Foundation, "Annual Report 2018", p31. Available from: https://www.iwf.org.uk/report/2018-annual-report

[26] NSPCC, "Taming the Wild West Web: How to regulate social networks and keep children safe from abuse", p6. Available from: https://www.nspcc.org.uk/globalassets/documents/news/taming-the-wild-west-web-regulate-social-networks.pdf.

[27] BBC, "Why games need to protect children from grooming". Available from: https://www.bbc.co.uk/news/uk-scotland-tayside-central-50226260.

[28] NSPCC, "Social isolation and the risk of child abuse during and after the coronavirus pandemic". Available from: https://learning.nspcc.org.uk/research-resources/2020/social-isolation-risk-child-abuse-during-and-after-coronavirus-pandemic.

[29] PrivSec Report, "TikTok under investigation following allegations over child data use". Available from: https://gdpr.report/news/2019/07/04/tiktok-under-investigation-following-allegations-over-child-data-use/.

for a search warrant or realistic prospects of conviction.[30]

Delays in investigating these types of cases have been identified in multiple police forces, including instances where suspects identified in indecent images of children cases were able to continue accessing children, including those in their own home.[31] A common factor in these failures is lack of skills and training. Research by Middlesex University in 2016 found that despite nearly two-thirds of surveyed officers having investigated online grooming or indecent images of children, the majority said that they had not received any relevant training and only 1 in 6 had received 'specific' training on the subject.[32] The resources to fight this type of crime are urgently need at a policing level – not a bureaucratic or regulatory one.

Adding an additional legal framework in this area poses a serious risk of detracting from the critical work already being carried out by regulators and industry alike to protect children from serious harm. Funding that could be better spent training police officers to investigate online crime involving children would instead go to more officials overseeing an areas that is already under the remit of other regulators. It would be a tragedy if the clamour to keep children safe online missed the real opportunity for change, leaving children more exposed than ever.

Cyberbullying

Cyberbullying and trolling – both of which are nasty but not actually illegal – are referenced in the White Paper, with statutory codes of practice published alongside it in line with Section 103 of the Digital Economy Act 2017. These are primarily aimed at ensuring that those who have suffered from cyberbullying are able to access the support they need.

> **“As indicated in UKCIS's name change, it is easy for the distinction between adults and children to be lost in an ever expanding remit.”**

The UK Council for Internet Safety (UKCIS, previously known as the UK Council for Child Internet Safety) already provides a range of guidance material on good practice for social media services. It brings together organisations with the specific focus on keeping children safe online, with more than 200 members drawn from across government, industry, law, academia and the charity sector. UKCIS was responsible, for example, for making it an unavoidable choice to ask home broadband customers whether they would like to turn on parental control filters.[33]

As indicated in UKCIS's name change, it is easy for the distinction between adults and children to be lost in an ever expanding remit. For a voluntary organisation, this is not a problem, but for a statutory regulator, it would be. This phenomenon can be seen in Australia's attempt to tackle online harms. A Children's eSafety Commissioner was established under the Enhancing Online Safety for Children Act 2015 to focus specifically on children's cyberbullying

[30] Home Affairs Select Committee, "Policing for the Future", p38. Available from: https://publications.parliament.uk/pa/cm201719/cmselect/cmhaff/515/515.pdf.

[31] Her Majesty's Inspectorate of Constabulary and Fire & Rescue Services, "National Child Protection Inspection Post-Inspection Quarter 4 Update". Available from: https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/metropolitan-national-child-protection-inspection-quarter-4-update.pdf.

[32] Middlesex University, "Enhancing Police and Industry Practice". Available from: https://www.mdx.ac.uk/__data/assets/pdf_file/0017/250163/ISEC-report-FINAL.pdf.

[33] GOV.UK, "UK Council for Child Internet Safety". Available from: https://www.gov.uk/government/groups/uk-council-for-child-internet-safety-ukccis

complaints.[34] The 'Children's' part of the title was quickly dropped as its role rapidly grew to cover safeguarding all Australians from online harm and promoting a safer and more positive online experience for adults as well as children.[35]

A new Online Safety Act is now being considered to formalise the Commissioner's expanded remit over adult cyberbullying, including over private messaging, and to harden the penalties for online abuse.[36] In just five years, what began as an exclusive complaints procedure for children's cyberbullying content that has not been removed from social media, has ballooned to an extensive suite of powers which may soon include takedown of legal private content in 24 hours, plus wide-ranging civil and criminal penalties for non-compliance.

In the UK White Paper, no parameters are given on whether separate codes would exist for adults rather than children, which has profound implications for freedom of expression. As discussed in section III, legal free speech content of this nature is heavily context-dependent – what may upset a nine-year-old may well be received as petty or humorous by a 30-year-old.

II: Protect national security

Alongside protecting children, preventing terrorism is at the top of the Government's agenda on online harms. Obvious parallels can be drawn between child sexual abuse content and terrorist content – both already illegal – and the most effective approach to regulation in this area.

Under the banner of terrorist content, there are important distinctions to be made between different types of content. For instance, it can involve communications for the direct planning of an attack, the promotion of propaganda from proscribed organisations, or the sharing of footage of an attack in progress.

Related concerns around national security, extremism and even public health have now been brought firmly into the narrative around harmful content in this area, where the lines of legality are greyer. According to the Government's own Prevent Strategy, speech critical of 'British values', such as democracy, counts as extremist. By this standard, social media companies could be judged ineffective at carrying out their duty of care if they fail to limit the distribution of alternative philosophical arguments. It would be left up to the regulator to decide whether or how far it wanted to take this.

Terrorist content

The legislation under which terrorism-related material is illegal was updated in the Counter-Terrorism and Border Security Act 2019, which includes an explicit section about online content. The terms of service of all leading social media companies stipulate that terrorist content is forbidden.[37]

As with CSAI, action to remove terrorist content functions through self-regulatory

[34] Australian Government eSafety Commissioner, "Children's eSafety Commissioner launches cyberbullying complaints scheme". Available from: https://www.esafety.gov.au/about-us/newsroom/childrens-esafety-commissioner-launches-cyberbullying-complaints-scheme.

[35] Stephen Lunn, The Australian, "It's an arms race against the worst of the worst". Available from: https://www.theaustralian.com.au/weekend-australian-magazine/esafety-commissioner-julie-inman-grant-and-the-battle-to-civilise-cyberspace/news-story/49ecb87c30ecdeaa87a6b12b5e157524.

[36] Australian Government Department of Infrastructure, Transport, Regional Development and Communication, "Consultation on a new Online Safety Act". Available from: https://www.communications.gov.au/have-your-say/consultation-new-online-safety-act.

[37] Stuart Macdonald, Sara Giro Correia and Amy-Louise Watkin, International Journal of Law in Context, 15(2), "Regulating terrorist content on social media: Automation and the rule of law". Available from: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/B54E339425753A66FECD1F592B9783A1/S1744552319000119a.pdf/regulating__terrorist__content__on__social__media__automation__and__the__rule__of__law.pdf.

industry initiatives run in cooperation with the police. For the last 10 years, the Counter-Terrorism Internet Referral Unit (CTIRU) has worked with social media companies to make referrals for removal of content. To date, in excess of 310,000 individual pieces of terrorist content referred by CTIRU have been removed by companies.[38] CTIRU also informed the design of the EU Internet Referral Unit based at Europol. Crucial to CTIRU's success is that it sits firmly within the Metropolitan Police, acting inside the confines of what is clearly prescribed by law as illegal.

> **“Crucial to CTIRU's success is that it sits firmly within the Metropolitan Police, acting inside the confines of what is clearly prescribed by law as illegal.”**

Following the Westminster terrorist attack in March 2017, the Government convened a roundtable with major industry players to look at how to reduce the availability of terrorist content online. Companies including Facebook, Twitter, Microsoft and Google came together to form the Global Internet Forum to Counter Terrorism (GIFCT) later that year, with other companies since having joined the consortium. The GIFCT has led the way in encouraging the development of automation and machine learning technology to detect and remove terrorist content, including a shared database of hashes (unique digital fingerprints) for content produced by or in support of terrorist organisations.

Hashes only help detect content or elements of content that can be matched with previous 'fingerprints'. Two terrorist attacks last year in Christchurch and Halle saw the perpetrators livestream footage of their assaults. Short of banning livestreaming altogether, there is precious little that an online harms regulator would be able to do to eliminate this possibility in the future.

In part due to the success of large social media platforms at automatically detecting and blocking terrorist content on their services, terrorists have largely been displaced onto hundreds of smaller sites. Some have been quicker to react than others, with sites like Telegram working with Europol to disrupt Daesh propaganda distribution via its encrypted services, conducting a mass purge of accounts.[39] More worryingly, there has also been a shift towards use of the dark net, which is inherently harder to police. For example, the day after the 2015 Paris attacks, Daesh's media arm, Al-Hayat Media Center, launched a new propaganda website on the dark net, including a video celebrating the attacks.[40]

While pushing terrorist content away from mainstream users is of course desirable, it does present new challenges for law enforcement, intelligence agencies and civil society groups working on the fringes of the internet.

[38] Cabinet Office, "Government Response to the Intelligence and Security Committee of Parliament Report 'Russia'". Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/902342/HMG_Russia_Response_web_accessible.pdf.

[39] Mubaraz Ahmed, Tony Blair Institute for Global Change, "After Christchurch: How Policymakers Can Respond to Online Extremism". Available from: https://institute.global/policy/after-christchurch-how-policymakers-can-respond-online-extremism.

[40] Steve Ragan, CSO News, "After Paris, ISIS moves propaganda machine to Darknet". Available from: https://www.csoonline.com/article/3004648/security-awareness/after-paris-isis-moves-propaganda-machine-to-darknet.html.

In terms of the future landscape, adding a quasi-policing role for an online harms regulator to this delicate mix cannot be the answer. An arms-length, independent regulator cannot and should not be privy to classified information about our national security. This area must be kept firmly in the hands of government and security services who can join the dots of the intelligence trail at the highest level.

Mis- and disinformation

The challenges of adjudicating complex and context-dependent legal speech has also been played out in platforms' attempts to combat mis- and disinformation. Here again, their own terms and conditions already go above and beyond what is prescribed by law, with many new tools being introduced over the last few years to carefully distinguish between legal expression that may be contrary to mainstream views and the sharing of deliberately misleading or fabricated content. The proactivity of platforms has been thrown into particularly sharp focus this year as false information and conspiracy theories about COVID-19 have spread over social media – and indeed with the recent publication of the Intelligence and Security Committee of Parliament report on Russia.

Mis- and disinformation content often contains elements of accurate information, mixed with misleading statements or false context, rather than being entirely fabricated. To add to this problem, mainstream news sites often report on celebrities' Tweets without clarification that their claims are unfounded and that they have already been flagged or removed for breaching community guidelines. Likewise, incidents of misinformation on broadcast media, including an 80-minute interview on London Live with David Icke on COVID-19 and 5G, have not been subject to any sanctions. Most social media firms are already going much further than mainstream news sites, broadcast media and the law in enforcement against repeat propagators of misinformation.

Well before the pandemic, platforms had been taking steps to make clearer what information may be disputed. Facebook had been tightening its fact-checking regime by changing its newsfeed algorithm to prioritise more trustworthy sources or more local news, and partnering with Full Fact to give more context on a claim's source. This system was enhanced for COVID-19 misinformation.[41] At the start of the year, Facebook also announced that it would appoint an independent oversight board to adjudicate on cases referred by the company itself or by users whom have exhausted the appeals process.[42]

Specific pop-ups have been brought in to combat misleading COVID-19 and anti-vaccination content, with groups and pages spreading misinformation being excluded from recommendations or predictions when you type in the search bar.[43] Similar search prompts have been introduced on Twitter for those searching for COVID-19 and 5G links. Google was a founding partner in First Draft, a non-profit network to expand and embed best practices in newsrooms and journalism schools around the world.[44] On YouTube, Google have added bio labels to identify state-funded news organisations.[45]

[41] Full Fact, "Full Fact to start checking Facebook content as third-party factchecking initiative reaches the UK". Available from: https://fullfact.org/blog/2019/jan/full-fact-start-checking-facebook-content-third-party-factchecking-initiative-reaches-uk/.

[42] Anthony Cuthbertson, The Independent, "Facebook reveals plans for Supreme Court-style oversight board". Available from: https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-oversight-board-supreme-court-mark-zuckerberg-a9109761.html.

[43] Facebook, "Combatting Vaccine Misinformation". Available from: https://about.fb.com/news/2019/03/combatting-vaccine-misinformation/.

[44] First Draft. Available from: https://firstdraftnews.org/about/.

[45] YouTube, "Greater transparency for users around news broadcasters". Available from: https://youtube.googleblog.com/2018/02/greater-transparency-for-users-around.html.

“At their core, all these remedies have in common that they are not actively banning legal expression, but using the tools at companies’ disposal to ensure other people are less exposed to dubious content by allowing users to make more informed judgements before sharing.”

As with so many other areas proposed to be brought into the scope of online harms regulation, this area is already awash with departmental and regulatory initiatives pulling in different directions. The Intelligence and Security Committee report on Russia highlighted the fragmentation caused by at least 10 different teams across government being involved in the overarching ‘Defending Democracy’ programme, even before the Electoral Commission and the ICO were taken into account.[46]

The latest addition to this clutter is the Counter Disinformation Unit, which appears to fill the role that the Government would be seeing an online harms regulator to provide – working with platforms to identify and remove harmful content that breaches their own terms and conditions.[47] It draws in expertise from across Whitehall, including teams from the Foreign and Commonwealth Office and military analysts from the Ministry of Defence.

To avoid direct duplication, this function would presumably be hived off to Ofcom as the new online harms regulator. However, for an independent arms-length body, the same level of access to classified security information is unlikely to be granted, meaning that important intelligence links could be missed. Here again, attempts to make us safer through an all-encompassing regulator could have the opposite effect to that which is desired.

III: Protecting freedom of expression

The top concerns highlighted in the Government’s initial response to the consultation White Paper were on the grounds of freedom of expression.[48] According to the Government’s analysis of consultees, concerns over freedom of expression were significantly more prevalent among individual respondents than organisations. Yet requests for greater protection for freedom of expression in the responses were largely brushed aside by the Government, with assurances that the regulator would act proportionately as long as companies had put in place reasonable processes to protect users. This offers little of the clarity that companies and users were hoping for on the practical safeguards against the regulator moving the goalposts of free speech.

According to Article 10 of the European Convention on Human Rights and the UK’s own Human Rights Act 1998, restrictions to free of expression have to be prescribed by law and necessary in a democratic society for a legitimate aim. New regulations to intervene against speech that is legal which impose sanctions on business activities as a result seem to fundamentally fail the ‘prescribed by law’

[46] Intelligence and Security Committee of Parliament, “Russia”, p12. Available from: http://isc.independent.gov.uk/committee-reports.

[47] Department for Digital, Culture, Media and Sport, Letter to Lord Putnam from Caroline Dineage MP. Available from: https://committees.parliament.uk/publications/1280/documents/11300/default/.

[48] Department for Digital, Culture, Media & Sport and Home Office, “Online Harms White Paper - Initial consultation response”. Available from: https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response.

test. For example, users must be able to foresee with reasonable certainty whether the platform will be legally obliged to remove content they are about to post.

> **“Given how alike the principles in the Loi Avia and the White Paper are, similar judicial challenges would likely be mounted in the UK on human rights grounds.”**

In June this year the French Constitutional Council found large parts of France’s online hate speech legislation (known as the Loi Avia after National Assembly member Laetitia Avia, who drafted it) unconstitutional. Among other reasons, it breached this legality test for impermissible vagueness. The intermediary’s obligation was not expressed in terms that enabled the scope of liability to be determined.[49] Given how alike the principles in the Loi Avia and the White Paper are, similar judicial challenges would likely be mounted in the UK on human rights grounds.

Loi Avia empowered France’s Higher Audiovisual Council to require hosts to remove the most extreme content (certain terrorist content and child sexual abuse imagery) within an hour. For other content, deadlines of up to 24 hours apply, depending on who has made the request, the nature of the content and whether the company is a host, platform or website publisher. This failed the requirement of necessity and proportionality, because determination of illegality was via the sole opinion of the administrative authority rather than the content itself, without the opportunity for the host to obtain a judicial ruling on the matter.

The UK Government’s proposals are likely to run aground on both of these aspects of legality too. If anything more so, as the White Paper expands much further into the terrain of ‘legal but harmful’ than the Loi Avia. The Foundation for Law, Justice and Society commented that the White Paper leaves open the possibility that constraints on free speech could be imposed ‘on the basis of opaque agreements between platforms and politician’ rather than being subject to the constraints of parliamentary debate.[50] If this fundamental principle is to change, it will require amendment or repeal of the Human Rights Act, requiring the full legislative scrutiny of Parliament. Even then, it could still be defeated under principles of freedom of speech under English common law. Extreme caution should be exercised in going down this route to tackle online harms so as not to erode important checks against creeping censorship.

Hateful and extremist content

Removing content on the grounds of it is hate speech relies on a particularly diffuse area of law with constantly changing precedents. Even though hate crime is placed on the White Paper’s list of harms with a ‘clear definition’, the reality is far murkier. For example, the conviction of Mark Meechan, who was fined after filming a pug performing a Nazi salute, highlighted that grossly offensive content with humorous intent falls under a form of hate crime law.

Hate crime law of course also covers far more serious acts, intended to stir up hatred against different groups, or physical harm involving hostility towards that group.

[49] Constitutional Council, “Decision 2020-801 DC of June 18 2020”. Available from: https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm.

[50] Damian Tambini, The Foundation for Law, Justice and Society, “Reducing Online Harms through a Differentiated Duty of Care”, p5. Available from: https://www.fljs.org/sites/www.fljs.org/files/publications/Reducing%20Online%20Harms%20through%20a%20Differentiated%20Duty%20of%20Care.pdf.

Different legal parameters of hate crime exist in UK law across the Public Order Act 1986, the Crime and Disorder Act 1998, the Criminal Justice Act 2003, the Malicious Communications Act 1988, the Racial and Religious Hatred Act 2006, the Communications Act 2003 (which Meechan was prosecuted under) and even the Football (Offences) Act 1991.[51]

Cases such as Meechan's have added to the ongoing debate over the balance between freedom of expression and offence in current hate crime laws in the UK. More recently, a man called Harry Miller was accused of transphobia for a series of Tweets criticising transgender orthodoxies, such as 'I was assigned mammal at birth, but my orientation is fish. Don't mis-species me.' He was visited by Humberside Police at work and told he had not committed a crime, but it counted as a 'hate incident'. The High Court found that the police had acted unlawfully, by breaching Miller's right to freedom of expression.

This episode further underlined the uncertainty around 'hate incidents' online. And definitions of hate crime may well change again in light of current campaigns to make misogyny a hate crime.[52] Scotland's Hate Crime Bill is grappling with the problem of making speech illegal simply if it is 'likely to stir up hatred', regardless of the intent behind the comments. It is hard to see how adding another level of legislation that regulates free speech according to its own measure of effectiveness will solve the problem.

The Meechan example, and many others besides, demonstrate that there are already laws against the plethora of different actions that can constitute hate or extremism. It is already possible to take effective enforcement action against the perpetrators of these crimes – by prosecuting those guilty of making the content in the first place. But incorporating imprecise and broad definitions of harm will increase the likelihood of firms erring on the side of caution when identifying and eliminating content that they may be sanctioned for: a 'takedown first, repeal later' approach that surely has grave implications for free speech.

> **"The Meechan example, and many others besides, demonstrate that there are already laws against the plethora of different actions that can constitute hate or extremism."**

This tendency will be exacerbated by the increasing use of AI to carry out such content-sifting. In practice, AI struggles to distinguish the nuanced nature of human expression which relies on understanding of culture, politics and most crucially, context. YouTube fell afoul of this when an update in its hate speech policy resulted in thousands of academic, journalistic and activist sites being removed.[53] The Index on Censorship points out examples such as documentation of war crimes being

[51] Mark Walters et al., University of Sussex, "Hate crime and the legal process". Available from: https://www.sussex.ac.uk/webteam/gateway/file.php?name=final-report---hate-crime-and-the-legal-process.pdf&site=539.

[52] Misogyny is included in Lord McNally's Private Members Bill: Online Harms Reduction Regulator (Report) Bill. Details available from: https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie__uk__trust/2019/12/17161822/Carnegie-UK-Trust-draft-ONLINE-HARMS-BILL.pdf.

[53] Julia Alexander, The Verge, "YouTube's new policies are catching educators, journalists, and activists in the crossfire". Available from: https://www.theverge.com/2019/6/7/18657112/youtube-hate-policies-educators-journalists-activists-crossfire-takedown-demonetization.

removed as 'hate speech', while anti-racist campaigners have found their sites removed for having used racial slurs as evidence of racism.[54]

Protecting marginalised groups

By pushing platforms to adopt a more zealous attitude to content removal, regulation may actually make matters worse for precisely those groups it is designed to protect. There is extensive evidence that certain groups in society are more likely to be victims of trolling and cyberbullying – women (especially young women), ethnic minorities, people with disabilities, LGBTQ+ people and those of particular faiths.[55]

Yet such marginalised groups are also disproportionately likely to be the victims of vexatious censorship attempts. Instagram has already displayed an over-zealous approach to 'shadow banning' queer and plus-sized bodies and women promoting their pole fitness businesses.[56] A US study showed that tweets created by African Americans are one and a half times more likely to be flagged as hate speech under leading AI models.[57] Internal documents from TikTok revealed that its moderators had been instructed to suppress (by removing from the 'For You' section) posts created by users deemed 'too ugly, poor, or disabled', including images of beer bellies, crooked smiles or even cracked walls.[58]

Any regulator that seeks to impinge on speech which may be deemed as 'offensive' could end up being used as a tool of censorship by vested interests. The battles by Mary Whitehouse's National Viewers' and Listeners' Association against bodies bound by statutory regulations, including the BBC, serve as a warning from the past on what can go wrong if regulators tend towards censorship. Encouraging more moderating – whether done by humans or machines – does not necessarily produce positive outcomes.

International implications

But we do not need to go back to the Whitehouse era to see harms regulation used to stifle political expression. Since around 2016, countries around the world have begun putting in place policies directly targeted at online platforms. Germany's Network Enforcement Act (NetzDG) mandates that social media companies delete what it calls 'manifestly unlawful' posts on their platforms within 24 hours of being notified or risk facing fines up to £44 million. Within a few months of its introduction, it had censored political parties, political satire and even the very politicians who pushed for its introduction.[59] More expansion of its powers are expected soon.

In recent years, a similar model of intermediary liability has been proposed

[54] Index on Censorship, "Submission to Online Harms White Paper consultation", p3. Available from: https://www.indexoncensorship.org/wp-content/uploads/2019/07/Online-Harms-Consultation-Response-Index-on-Censorship.pdf.

[55] Centre for Strategy & Evaluation Services, "Rapid Evidence Assessment: The Prevalence and Impact of Online Trolling", p16-18. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/811449/DCMS_REA_Online_trolling_.pdf.

[56] Chanté Joseph, The Guardian, "Instagram's murky 'shadow bans' just serve to censor marginalised communities". Available from: https://www.theguardian.com/commentisfree/2019/nov/08/instagram-shadow-bans-marginalised-communities-queer-plus-sized-bodies-sexually-suggestive.

[57] Sap et al., Association for Computational Linguistics, "The Risk of Racial Bias in Hate Speech Detection". Available from: https://www.aclweb.org/anthology/P19-1163/.

[58] Sam Biddle, Paulo Victor Ribeiro, Tatiana Dias, The Intercept, "Invisible Censorship: TikTok told moderators to suppress posts by "ugly" people and the poor to attract new users". Available from: https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/.

[59] For example, see Linda Kinstler, The Atlantic, "Germany's Attempt to Fix Facebook is Backfiring". Available from: https://www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/.

or adopted in at least 10 countries which are classed as 'not free' or 'partly free' in Freedom House's 2019 assessment of freedom of the internet.[60] Several of these countries now require internet intermediaries to remove content falling under broad interpretations of 'defamation of religions', 'anti-government propaganda' and even simply 'unreliable information'. The latter category, which has been adopted in Russian law, has been defended by the Kremlin by explicitly citing NetzDG as an example of false information being 'regulated fairly harshly' in other European countries.[61] Similar rules on 'false statements of fact' in Singapore's Online Falsehoods and Manipulation Bill look like they will be used to target human rights and civil society groups criticising government.[62]

> **"In recent years, a similar model of intermediary liability has been proposed or adopted in at least 10 countries which are classed as 'not free' or 'partly free' in Freedom House's 2019 assessment of freedom of the internet."**

In developing countries, crackdowns on internet free speech can have devastating consequences for internet access for their citizens. In 2018 Tanzania brought in requirements that online content creators pay the rough equivalent of £770 in registration and licensing fees.[63] Contributors' details must be stored for 12 months, financial sponsors disclosed and cyber cafes must have surveillance cameras. Once online, you could be hit with a £1,750 fine and/or jail time if your content is deemed 'indecent', 'leads to public disorder' or is simply 'annoying'.[64] Meanwhile, in neighbouring Uganda, a new 'Over-The-Top' social media tax has been introduced to curtail 'gossip'.[65] Beyond the dystopian implications, these measures have a deeper impact on the poorest in their societies by effectively pricing them out of the internet.

It is not just in the UK where freedom of speech is at serious risk from overreach into the greyer areas of legal expression. If as a country we are seeking to lead the world in internet safety, this must include considering how the precedent set by our regulation may be used by governments less concerned with the fundamental freedoms of their citizens.

[60] Jacob McHangama, Foreign Policy, "Germany's Online Crackdowns Inspire the World's Dictators". Available from: https://foreignpolicy.com/2019/11/06/germany-online-crackdowns-inspired-the-worlds-dictators-russia-venezuela-india/.

[61] Maria Vasilyeva and Tom Balmforth, Reuters, "Russia's parliament backs new fines for insulting the state online". Available from: https://www.reuters.com/article/us-russia-politics-fakenews/russias-parliament-backs-new-fines-for-insulting-the-state-online-idUSKBN1QU1UN.

[62] Jacob McHangama, Foreign Policy, "Germany's Online Crackdowns Inspire the World's Dictators". Available from: https://foreignpolicy.com/2019/11/06/germany-online-crackdowns-inspired-the-worlds-dictators-russia-venezuela-india/.

[63] Shayera Dark, The Verge, "Strict new internet laws in Tanzania are driving bloggers and content creators offline". Available from: https://www.theverge.com/2018/7/6/17536686/tanzania-internet-laws-censorship-uganda-social-media-tax.

[64] Ibid.

[65] Deutsche Welle, "Uganda: One year of social media tax". Available from: https://www.dw.com/en/uganda-one-year-of-social-media-tax/a-49672632

IV: Promoting business

The story of technology businesses in the UK is an overwhelmingly positive one. In 2018, 7.7% of total UK Gross Value Added came from the digital sector alone.[66] Before the COVID-19 crisis, it was growing six times faster than the wider economy, adding £400m to the UK economy every day.[67] A recent Centre for Policy Studies (CPS) report, 'Platforms for Growth', highlighted the transformative effect that platforms have on the performance and productivity of businesses in Britain, especially small- and medium-sized ones.[68] Big tech serves as a boost to the rest of the UK economy.

Another previous report by the CPS, 'Herding Unicorns', showed how Britain punches above its weight in its number of 'unicorn' companies[69] – unlisted tech companies with valuations of more than £1 billion. The latest estimate from Beauhurst now puts the UK unicorn tally at 17.[70] In short, the UK has a flourishing technology ecosystem, built on a business environment that promotes entrepreneurship, profit-making and job creation.

The costs that regulation imposes on a sector have a direct impact on the shape of the market. Increasing fixed costs through compliance burdens will inevitably reduce the dynamism of the UK technology ecosystem. That will in turn give rise to market ossification, whereby large incumbents are allowed to dominate, which not only reduces consumer welfare but also limits the opportunities for private individuals to participate in the economy.

> **"Increasing fixed costs through compliance burdens will inevitably reduce the dynamism of the UK technology ecosystem."**

On this front, not only is the White Paper wide in the scope of the harms it includes, but also in the array of different online operators over which it proposes that the new regulator will have jurisdiction. The regulatory framework in the White Paper is designed to apply to all companies no matter their size that 'allow users to share or discover user-generated content or interact with each other online'.[71] It is not difficult to imagine what a vast swathe of the economy and society that covers, from small businesses that allow customers to leave reviews of their content, to video games which allow players to exchange messages or actions, to Instagram entrepreneurs, to firms running Twitter accounts recycling the latest football news or humour.

Although the Government's response to the White Paper has attempted to narrow the types of services in scope, such as removing business-to-business (B2B)

[66] Department for Digital, Culture, Media and Sport, "DCMS Sector Economic Estimates 2018 (provisional): Gross Value Added". Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/863632/DCMS_Sectors_Economic_Estimates_GVA_2018.pdf.

[67] Department for Digital, Culture, Media and Sport, "Digital sector worth more than £400 million a day to UK economy". Available from: https://www.gov.uk/government/news/digital-sector-worth-more-than-400-million-a-day-to-uk-economy.

[68] Eamonn Ives, Centre for Policy Studies, "Platforms for Growth". Available from: https://www.cps.org.uk/research/platforms-for-growth/.

[69] Centre for Policy Studies, "Herding Unicorns: How Britain can create and support the high-growth tech companies of the future". Available from: https://www.cps.org.uk/files/reports/original/190301101443-CPSHerdingUnicornsFINAL.pdf.

[70] Beauhurst, "UK Startup Unicorns: The complete List 2020". Available from: https://about.beauhurst.com/blog/uk-startup-unicorns/.

[71] HM Government, "Online Harms White Paper", p8. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

services, many online business models do not fit neatly into the inclusion or exclusion terms. Their assessment is that fewer than 5% of UK business would fit into scope – but that still leaves nearly 300,000 firms. Understandably, this has been a cause for alarm among businesses seeking clarity on their regulatory status.[72]

Investment

In 2019, more than £10 billion was invested in UK tech, the most for any single country after the USA and China.[73] Of the top 20 cities for tech investment in Europe, the UK boasts five – including the fastest-growing tech cluster in Europe, Manchester, which saw investment rise from £48m in 2018 to £181m in 2019, a rise of 277%.[74]

> **“Investors seek both stability and consistency, as well as proportionate and predictable regulation, when deciding where to put their money for the long term.”**

Unclear and overbearing regulation could have stark consequences for investment in the sector. A survey of investors in the tech sectors in the UK, Ireland, France and Germany showed remarkable similarities in their views on the challenges for the industry going forward, particularly for start-ups and scale-ups.[75] Investors seek both stability and consistency, as well as proportionate and predictable regulation, when deciding where to put their money for the long term.

Most striking in the UK responses, summarised in the chart below, was the strength of feeling around the disastrous unintended consequences for start-ups of designing policy around the current dominant players – which is overwhelmingly what has happened when it comes to online harms regulation. More than two-thirds agreed that changes to liability could make them reassess whether to invest in UK platform businesses. In addition, plans to fund the regulator through an additional digital levy could further add to the disproportionate burden on smaller and newer companies.

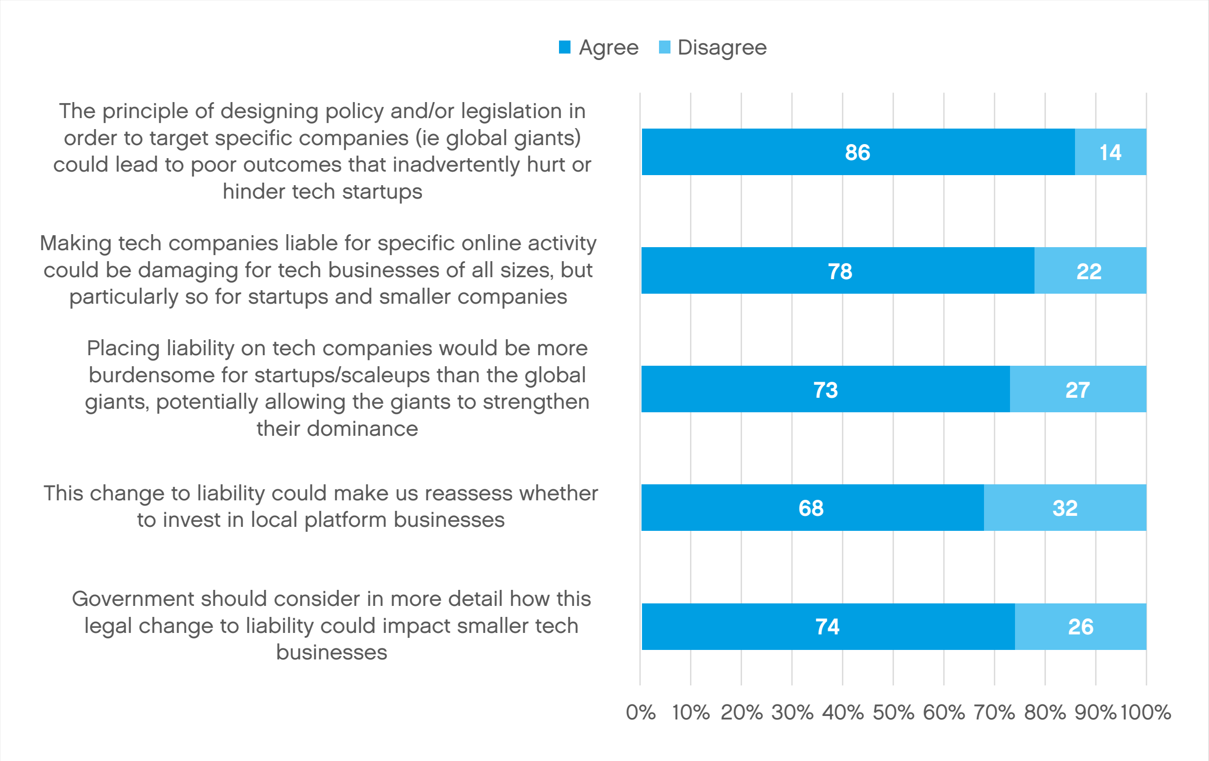
[72] Department for Digital, Culture, Media & Sport and Home Office, “Online Harms White Paper - Initial consultation response”. Available from: https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response.

[73] Tech Nation, “UK tech for a changing world”. Available from: https://technation.io/wp-content/uploads/2020/03/Tech-Nation-Report-2020-UK-Tech-For-A-Changing-World-v1__0.pdf.

[74] Eamonn Ives, Centre for Policy Studies, “Platforms for Growth”. Available from: https://www.cps.org.uk/research/platforms-for-growth/.

[75] Coadec, “The Impact of Regulation on the Tech Sector”, p6. Available from: http://coadec.com/wp-content/uploads/2018/12/The-Impact-of-Regulation-on-the-Tech-Sector.pdf.

Investor views on the impact of regulation on the tech sector



Source: The Coalition for a Digital Economy (Coadec) in partnership with European partners Allied for Startups, France Digitale & Silicon Allee[76]

The evidence of lost investor confidence from regulation is not just hypothetical – the introduction of the GDPR has been estimated to result in a yearly loss of up to 29,000 jobs in the EU.[77] What its creators hoped would foster a more competitive and innovate digital market in fact looks to have increased concentration, principally to the benefit of Facebook and Google.[78]

Compliance burdens

Small businesses are already disadvantaged by the burden of regulation in the UK.

As our previous CPS report, 'Think Small' highlights, regulatory compliance takes up a disproportionally large share of small businesses' time and resources. The Department for Business, Energy and Industrial Strategy's own Business Perceptions report shows that there is an 'upward trend' of businesses being concerned by regulatory compliance.[79] The mean number of days spent dealing with regulation is 5.1 days for micro-firms (those with between one and nine employees) and 8.7 days for small companies (10-49

[76] Ibid.

[77] Victoria Hewson and James Tumbridge, IEA, "Who Regulates the Regulators? No.1: The Information Commissioner's Office", p15. Available from: https://iea.org.uk/wp-content/uploads/2020/07/Who-regulates-the-regulators__.pdf.

[78] Garrett Johnson, Scott Shriver and Samuel Goldberg, SSRN, "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR". Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract__id=3477686

[79] Department for Business, Energy and Industrial Strategy, "Business Perceptions Survey 2018", p15. Available from: https://www.gov.uk/government/publications/business-regulation-business-perceptions-survey-2018.

employees).[80] This impact can also be seen in productivity per worker in different sized firms. For micro-firms, the median average productivity per worker is £24,000, compared to £37,000 for large firms (250-999 employees).[81]

“As with GDPR, wide-ranging regulations such as those outlined in the White Paper would be difficult for many small companies to fulfil in-house, leading to expensive bills for outside advice and assistance.”

As with GDPR, wide-ranging regulations such as those outlined in the White Paper would be difficult for many small companies to fulfil in-house, leading to expensive bills for outside advice and assistance. The average number of days spent dealing with regulation is a number that a Conservative government ought to be concerned with driving downwards, and it is through this lens that the proposed regulation ought to be scrutinised.

Barriers to competition

Size really matters when considering the burden of compliance and enforcement of new regulations. Currently, there is no minimum size threshold for triggering the White Paper's regulations – all firms in scope must comply from day one. Discretion is left up to the regulator to decide the appropriateness of its enforcement action, with ‘due regard for innovation’.

Among the top arguments used by those advocating more stringent platform regulation is that it has been proven to force companies to step up their moderating functions. Germany is often lauded for ‘taming’ Facebook through the requirements of NetzDG, which led to the vast expansion of ‘deletion centres’ in Germany. In its third party moderator's offices in Berlin there are over 1,200 moderators and four trauma specialists.[82] Across the world, Facebook has more than 30,000 employees working on safety and security — about half of whom are content moderators – and there have been recent calls for moderator numbers to double.[83] YouTube and Twitter have likewise increased their moderating capacity over the last few years, often needing to outsource these jobs abroad where labour costs are cheaper.[84]

Employing human moderators comes at a huge cost, but for social media giants, it is a relatively tolerable cost. For smaller companies looking to scale up, however, it is a huge barrier to their expansion, as the monitoring and filtering obligations will represent a much higher proportion of their costs than for more established players. The Foundation for Law, Justice and Society summarise the seriousness of the situation that the White Paper may cause:

[80] Ibid.

[81] Office for National Statistics, “Understanding firms in the bottom 10% of the labour productivity distribution in Great Britain: “the laggards”, 2003 to 2015”. Available from: https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/articles/understandingfirmsinthebottom10ofthelabourproductivitydistributioningreatbritain/jantomar2017#results.

[82] Katrin Bennhold, The New York Times, “Germany Acts to Tame Facebook, Learning From Its Own History of Hate”. Available from: https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html.

[83] Charlotte Jee, MIT Technology Review, “Facebook needs 30,000 of its own content moderators, says a new report”. Available from: https://www.technologyreview.com/2020/06/08/1002894/facebook-needs-30000-of-its-own-content-moderators-says-a-new-report/.

[84] Elizabeth Dwoskin, Jeanne Whalen and Regine Cabato, The Washington Post, “Content moderators at YouTube, Facebook and Twitter see the worst of the web — and suffer silently”. Available from: https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/.

> 'One of the most likely scenarios is that, as a result of the new duty of care, the costs of compliance may become so great that smaller or medium-sized companies, if within scope, may decide that rather than comply they will simply cease to provide services that fall within the definition of scope for the White Paper. A very real danger is that YouTube and Facebook are in fact the only services with the resources to provide the necessary levels of moderation.'[85]

Even if the regulator choses a light touch approach in sanctioning small businesses, the risk-reward calculation for entrepreneurs will be less favourable if they expect arbitrary and harsh compliance standards when they start to scale up. For innovative new firms, the challenge of competing in an already tough marketplace will be greater than ever.

On enforcement too, inconsistency further penalises smaller firms. Large companies that have been subject to fines by the ICO have often been able to use their considerable financial and legal resources to challenge these decisions. Last year Facebook appealed against a fine of £500,000 on the grounds of bias and procedural unfairness. The ICO was forced to settle this out of court, allowing Facebook to avoid admitting liability (although they did still have to pay the fine).[86] Much larger cases against the likes of British Airways and Marriott Hotels have been delayed as it appears the ICO is unable or unwilling to defend its position against the legal firepower large corporations are able to bring to the fight.[87]

> **“Even if the regulator choses a light touch approach in sanctioning small businesses, the risk-reward calculation for entrepreneurs will be less favourable if they expect arbitrary and harsh compliance standards when they start to scale up.”**

Clearly smaller players would not have such overwhelming resources to wield over the new online harms regulator to protect themselves from unfair applications of the rules. Meanwhile, transparency on where the boundaries lie is unable to filter down from larger cases which go to settlement, causing even more uncertainty for those lower down the pecking order.

It is easy to forget that the likes of Facebook and Twitter are still only 16 and 14 years old respectively. TikTok is just three. Those proposing new regulations should reflect on whether they intend to prevent the emergence of the next generation of these types of companies in the UK by making it near-impossible to succeed here.

[85] Damian Tambini, The Foundation for Law, Justice and Society, “Reducing Online Harms through a Differentiated Duty of Care”, p3. Available from: https://www.fljs.org/sites/www.fljs.org/files/publications/Reducing%20Online%20Harms%20through%20a%20Differentiated%20Duty%20of%20Care.pdf.

[86] First Tier Tribunal, General Regulatory Chamber (Information Rights), “Preliminary Issue Ruling”. Available from: https://panopticonblog.com/wp-content/uploads/sites/2/2019/07/033-270619-Preliminary-Issue-Ruling-Facebook-Ireland-and-IncEA20180256.pdf.

[87] Victoria Hewson and James Tumbridge, Institute for Economic Affairs, “Who Regulates the Regulators? No.1: The Information Commissioner's Office”, p23. Available from: https://iea.org.uk/wp-content/uploads/2020/07/Who-regulates-the-regulators__.pdf.

Part 3: An alternative model

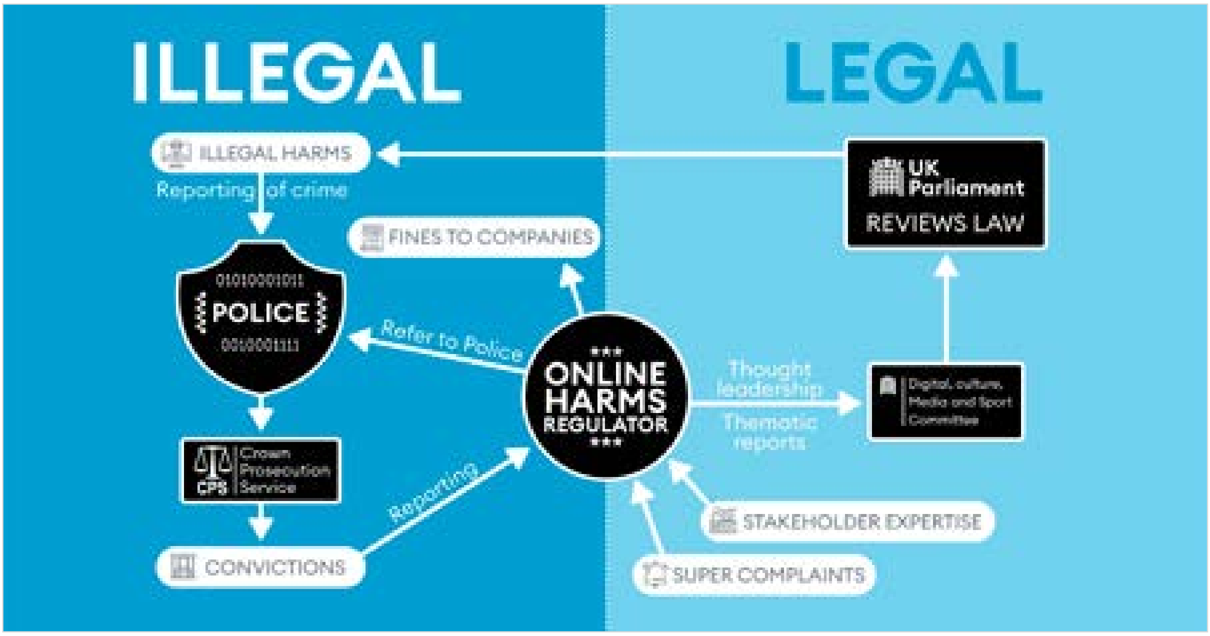
This report has highlighted the many grievous problems with the current proposals for online harms regulation – not least the fact that it is likely to be declared illegal. So what should be done instead?

We propose that Ofcom takes responsibility for online harms regulation, but with entirely separate functions for dealing with illegal and legal content. Ofcom's powers should only allow it to take enforcement action where a particular platform has not acted to improve despite specific court cases relating to illegal content on its site.

For content that is presently legal, Parliament should decide whether it is sufficiently harmful that the state should suppress it. This is a matter for lawmakers to specify by giving it criminal status, not for Ofcom to work out later. If the Government is not prepared to criminalise certain speech, then it should not punish social media companies for giving it a platform: if something is legal to say, it should be legal to type. Ofcom should feed into the process of recommending when the law should be changed by processing information from external sources in reports to Parliament.

The police are ill equipped to deal with the nature and magnitude of internet crime. By limiting the regulator's role, precious time, money and resources can instead be diverted to stamping out genuinely illegal content at its source – the user. This would tackle online harm far more effectively, without the risk of it coming at the cost of freedom of expression, or impacting on the thriving technology sector in the UK.

Legal vs Illegal



Illegal harms

1. Illegal content should be reported directly to the police. If illegal content is reported to the regulator, this should be referred immediately to the police.
2. The police should investigate the crime in the normal way with the Crown Prosecution Service deciding whether to charge.
3. If a conviction on illegal content is handed down, Ofcom should add this to its reports on other similar convictions.
4. Where evidence shows that companies are repeatedly failing to act to improve processes, Ofcom would have the power to investigate, request improvements and fine that company.

Legal harms

1. For legal harms, Ofcom would have a research and thought leadership function on these issues from industry, civil society organisations, academic research, police and 'super-complaint' groups.
2. Ofcom would compile the evidence for Parliament to scrutinise.
3. Parliament would debate whether individual content or practices should be made illegal.
4. Once a harm was illegal, it would be dealt with by the illegal side of Ofcom's remit on online harms.

Example case: Encouragement to self-harm

There has been a high level of media attention over the past few years about advocacy of self-harm on social media platforms. The Government might want to consider whether it is time for a law covering the offline and online worlds regarding encouraging self-harm.

In the White Paper, suicide and self-harm are dealt with together. However, encouraging someone to commit suicide is illegal, while encouraging someone to self- harm is not.

As with many other online harms, self-harm content is highly subjective and outright censorship of the topic would be extremely dangerous. Careful judgements need to be made to separate out legitimate discussion from coercive encouragement to commit self-harm. For instance, supportive channels discussing survivors' stories may include graphic descriptions that are triggering to some but be immensely helpful to others in sharing ways to cope and recover. Likewise, definitions of what constitutes an act of self-harm are not universally agreed upon either.

Through our model of regulation, evidence could be gathered by Ofcom from court cases, social media platforms themselves, academic studies, charities and victims' groups on encouragement to self-harm. Ofcom could then make recommendations to put before the Department for Digital Culture Media and Sport (DCMS) Select Committee on how encouragement to self-harm manifests itself online. The Committee could compare this with evidence from the offline world to make sure that any policy recommendations put forward to Parliament to pass legislation dealt with the issue in the round.

By subjecting any rule changes to full parliamentary scrutiny, this model guards against the regulator or the government ratcheting up its powers easily. Under the White Paper, it was intended that the Government would be given a direct say over writing certain codes on child sexual abuse and terrorism, effectively giving it a back door into the regulator if officials could argue it was for national security reasons.

> **“By subjecting any rule changes to full parliamentary scrutiny, this model guards against the regulator or the government ratcheting up its powers easily.”**

For example, a case could be made about the threat of foreign interference in our elections being a threat to national security, allowing government to make the case for directly writing the codes of practice on disinformation. Structuring regulation on this basis violates the most basic principles of independence and proportionality on which all good regulation should be built. The principle of democratic scrutiny must stand.

Scope

Ofcom's scope should cover services whose primary function is public sharing of user-generated content. It should exclude private messaging, comment or review sections, B2B services and any other services until or unless it is directed to by government. Avoidance would therefore be less problematic than under the White Paper regime given the less onerous compliance that this requires. It also ensures that online service providers are less likely to deliberately distort their activities in order to avoid falling in scope.

We recommend that all services in scope should display the Ofcom mark to indicate that they have made themselves accountable to the regulation it produces and as a sign of commitment to high standards of internet safety. In return, their terms and conditions should be expected to reflect the best practice coming out of Ofcom's evidence-gathering function. Ofcom should also provide signposting on the difference between legal and illegal harms, with instructions about reporting illegal content to the police and the relevant avenues available to users to raise individual complaints through the platforms themselves.

Private communication

No firm decision appears to have been made in the Government's initial response as to the inclusion of private communications within the scope of the new online harms regulation. While it is acknowledged that some of the most serious illegal activity occurs through private messaging, the privacy rules on monitoring these spaces are tight. The Investigatory Powers Act 2016 already contains provisions to enable the lawful interception and obtaining of communication data. It allows for police and security services to apply for warrants to intercept communication data only when it is necessary and proportionate on statutory grounds of national security or serious crime.[88] The Information

[88] Home Office, “Interception of Communications: Code of Practice”. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/496064/53659_CoP_Communications_Accessible.pdf.

Commissioner's own response to the White Paper comments that the checks and balances in this Act reflect the importance of supporting the ability of the vast majority of people to enjoy private and secure communications in their everyday life.[89]

> **"If private communication remains in scope of the online harms regulation then it is hard to see how a regulator would be able to apply the White Paper's codes for this area fairly without creating a multi-tier system for private channels"**

Many of the most popular private messaging services use end-to-end encryption or are considering switching to it as a way of ensuring the security and privacy that their users demand.[90] The platforms themselves do not have the ability to read their users' encrypted messages, making monitoring and regulating impossible. There have been international calls from the likes of the Five Eyes alliance for backdoors to be built into encrypted services – but these have been undermined by their own practices of sharing intelligence across borders, crossing multiple legal and ethical boundaries.

Protection from the prying eyes of businesses and governments is precisely why so many, not least politicians themselves, choose to use encrypted services for communicating. If private communication remains in scope of the online harms regulation then it is hard to see how a regulator would be able to apply the White Paper's codes for this area fairly without creating a multi-tier system for private channels. Texts and emails would be out of scope; other non-encrypted messaging would be in scope; and encrypted messaging would be in scope but un-enforceable. It could even have the opposite effect the Government intends, pushing towards encryption as the norm for private communication in order to avoid the compliance burden.

Business services

The Government has acknowledged that B2B services should be out of scope. This is welcome, particularly for business operating in areas like on-demand cloud computing. However, there are a number of services used by businesses that do not necessarily fit neatly into the B2B box. For example, academic collaboration sites which allow people to share and comment on research would presumably still be in scope. Similarly, enterprise software such as Slack which offers increasingly integrated functions for communication within and between companies is likely to contain private and commercially sensitive information that clients would be uncomfortable with being actively monitored. It is also exactly these sorts of services which create the optimum conditions for innovative new ideas to be developed in the UK technology sector.

Likewise, the Government has stated that companies using social media in their business activities would not be in scope, as the platforms themselves would be responsible. However, online businesses

[89] Information Commissioner's Office, "The Information Commissioner's response to the Department for Digital, Culture, Media & Sport consultation on the Online Harms White Paper", p13. Available from: https://ico.org.uk/media/about-the-ico/consultation-responses/2019/2615232/ico-response-online-harms-20190701.pdf.

[90] For example, see Mark Zuckerberg, Facebook, "A Privacy-Focussed Vision for Social Networking". Available from: https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/.

are increasingly integrating 'social' features into their websites. For instance, live chat functions are an increasingly popular way for customers to get in touch with businesses and for businesses to triage issues through partial automation. Likewise, product reviews, which are a hugely valuable tool for consumers to distinguish between products and sellers, would still fall in scope.

> **"It is equally easy to see how a large company using an innovative start-up for its business services could pass the buck onto them if regulatory enforcement was threatened."**

These services are highly unlikely to be an avenue for online harm, but nonetheless involve user interactions through content that would need to comply with new regulation under the White Paper's plans. A scenario in which a super-complaint is lodged against a business service on the grounds of cyberbullying or hate crime on a private business channel or via a comments section is easy to envisage. The company using the business service may well have different terms and conditions to the company it is being used by. It is equally easy to see how a large company using an innovative start-up for its business services could pass the buck onto them if regulatory enforcement was threatened.

Leaving space for the law to adapt to emerging practices while being clear about which services are in scope is a difficult balance. This is why it must be the policy that private communication and business services are excluded from the outset. If a certain type of business service later emerges that cross into the social media space, then this should require Parliament to explicitly approve the addition of this service, not leave it up to the regulator to pick its own target.

Liability

The international nature of online platforms makes liability a particularly thorny issue for UK regulators. Most of the online platforms that would be in scope are headquartered outside the UK, predominantly in the US or China. Recent analysis of eight official documents released over the last two years showed that over three quarters of the references to specific firms were about just two US companies (and their subsidiaries): Google and Facebook.[91] Chinese firms made up just 1% of references and only two platforms in Europe were mentioned at all (Spotify and Ecosia). Not a single UK firm featured. Any sanctions regime must take account of the fact that the bulk of companies and their directors it would be seeking to take enforcement action against are headquartered overseas.

Platform liability

A key distinction between the White Paper and the model outlined here is the balance of liability. At present under the EU e-Commerce Directive's (ECD) safe harbour provision, online operators are not liable for illegal content that is hosted on their services unless and until they have notice of it. They must act expeditiously to remove or disable access to the illegal content once it is brought to their attention.

This type of liability regulation of taking 'reasonable' steps to keep people safe through a duty of care exists in other areas of UK law, such as Occupier Liability Act 1957, the Health and Safety Act 1974 and the

[91] Philip Schlesinger, London School of Economics, "The Changing Shape of Platform Regulation". Available from: https://blogs.lse.ac.uk/medialse/2020/02/18/the-changing-shape-of-platform-regulation/.

Environmental Protection Act 1990.[92] The core problem with this far-reaching duty as set out in the White Paper is that in practice, taking reasonable steps would amount to an obligation to monitor activity on a vast scale and in a highly nuanced way. It is also likely to be illegal under several European Commission Directives, which have been adopted under UK law. General monitoring and filtering systems to target specific types of content while indiscriminately monitoring all information shared by platform users for an unlimited period of time would not at present meet 'the requirement that a fair balance be struck between the right to intellectual property, on the one hand and the freedom to conduct business, the right to protection of personal data and the freedom to receive or impart information, on the other'.[93]

> **“The core problem with this far-reaching duty as set out in the White Paper is that in practice, taking reasonable steps would amount to an obligation to monitor activity on a vast scale and in a highly nuanced way.”**

The EU is currently consulting on a new Digital Services Act which could alter the safe harbour provision. It follows court cases including Eva Glawischnig-Piesczek vs Facebook (Case C-18/18) which have begun to erode the principle that intermediaries have no general obligation to monitor content.[94] However, progress is faltering in the face of opposition led by Ireland, Finland and Sweden to a planned 3% levy on companies earning €750 million in revenue, €50 million of which would need to be EU taxable revenue.[95]

Even so, the EU proposals so far look far less radical than those in the White Paper, which should ring alarm bells on Whitehall.

The only relevant directive it has so far ruled out implementing after Brexit is the EU Copyright Directive. Article 17 of this directive (previously Article 13) – dubbed the 'meme ban' – requires that online platforms stop copyrighted content getting onto their platforms.[96] This would make some platforms liable for content uploaded by their users, albeit not on the grounds of harm. Such liability would mean platforms needing widespread automated filters for non-harmful content, which could have serious unintended consequences for user-led innovation.[97]

Potential changes to Section 230 liability in the United States are also relevant given the number and size of the platforms based in the US. Section 230 was enacted in 1996 as part of the Communications Decency Act, allowing an 'interactive computer service' to take down content that is offensive or otherwise objectionable as long as they act in 'good faith' – very similar to the 'safe harbour' provision later included in the ECD. Most of the Act was struck down by the

[92] International Institute of Communications, “Duty of Care”, p18. Available from: https://www.iicom.org/images/iic/intermedia/jan-2019/im2019-v46-i4-dutyofcare.pdf.

[93] Privacy International, “Privacy International's Response to the Open Consultation on the Online Harms White Paper”, p3. Available from: https://privacyinternational.org/sites/default/files/2019-07/Online%20Harms%20Response%20-%20Privacy%20International__0.pdf.

[94] Corryne McSherry, Electronic Frontier Foundation, “Bad News From the EU: High Court Blesses Global Takedown Order”. Available from: https://www.eff.org/deeplinks/2019/10/bad-news-eu-high-court-blesses-global-takedown-order.

[95] Beatriz Rios and Samuel Stolton, Euractiv, “Parliament threatens to withhold consent on budget cuts, potentially delaying recovery plans”. Available from: https://www.euractiv.com/section/economy-jobs/news/parliament-threatens-to-withhold-consent-on-budget-cuts-potentially-delaying-recovery-plans/.

[96] Written Question, “Copyright: EU Action” answered by Chris Skidmore. Available from: “https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2020-01-16/4371.

[97] CREATe, “EU Copyright Reform”. Available from: https://www.create.ac.uk/policy-responses/eu-copyright-reform/.

courts as an unconstitutional infringement on free speech, but Section 230 remains.

As part of US-UK free trade negotiations, the UK has made clear that any deal must 'ensure the Government maintains its ability to protect users from emerging online harms'.[98] Platforms have warned that some of the changes mooted would force them to take a much more hands-off approach to moderating content. As Jack Dorsey of Twitter put it: 'If we didn't have that protection, we would not be able to do anything around harassment or to improve the safety or health of the conversation around the platform.'[99] Maintaining flexibility for the UK to diverge from future possible reforms in the US is therefore crucial.

Criminal Liability for Directors

A further step under consideration is the imposition of criminal liability on the senior management of online platforms. This would be a mistake and it will not be replicated in our alternative model for both reasons of legality and practicality. Criminal regulatory offences arising from company directorships are not completely without precedent, but must follow basic principles of duty under the law, including foreseeability, proximity, fairness, justice and reasonableness. The Government's proposed plan fails on these grounds. Legally speaking, personal liability would 'ordinarily require that he procured, directed, authorised a commission of the tort in question', or *mens rea*.[100] The current regulatory framework of the digital services sector, and in particular social media platforms, is not at the scale or specificity of the other sectors where criminal liabilities have been imposed on company directors.

Lawfulness, fairness and transparency are the cornerstone of all good regulation. Indeed, they are the stated first principles of a similar regulatory area: the GDPR. All three of these are under threat by introducing liability of this nature. To avoid the situation above the regulator could, of course, use its discretion not to pursue action against hostile foreign powers, but this risks creating an obvious loophole for companies and their directors seeking to operate at the margins of what is acceptable. Even more worryingly, it could drive away legitimate businesses with senior leadership teams who simply do not want to risk falling foul of vague UK rules if there are problems on their platform.

> **“Lawfulness, fairness and transparency are the cornerstone of all good regulation…, All three of these are under threat by introducing liability of this nature.”**

Independence

Independence is the most basic founding principle of any regulator. For Ofcom, an already vast institution, adding a new remit over our online activities would make it a hugely powerful body, as discussed above. With this in mind, it is more important than ever that its online harms functions are organisationally entirely separate from its other responsibilities and that its role remains tightly limited.

Ofcom's online harms regulation division should be made up of an independent

[98] James Titcomb, The Telegraph, “The internet law that could be central to a US-UK trade deal”. Available from: https://www.telegraph.co.uk/technology/2020/06/14/internet-law-could-central-us-uk-trade-deal/.

[99] Ashley Gold and Joanna Plucinska, Politico, “US, Europe threaten tech industry's cherished legal 'shield'”. Available from: https://www.politico.eu/article/tech-platforms-copyright-e-commerce-us-europe-threaten-tech-industrys-cherished-legal-shield/.

[100] Mr Justice Scott Baker in Great Western Trains Unreported (30 June 1999) - see also R v P [2007] EWCA Crim 1937 and Wilson v R [2013] EWCA Crim 1780.

oversight board, with a Chair and CEO hired by an appointments panel. The Board should monitor performance, provide advice, challenge issues and support the strategic direction specific to this part of Ofcom. Some board members would have no connection to Ofcom or the online sector (lay members), while others would have senior experience and offer expertise in online standards. The Board should be responsible for appointing the Reporting Committee, but be separate from its decision-making.

> **“Ofcom’s online harms regulation should be underpinned by principles of international human rights law, using the established UN human rights framework to set out public interest objectives for online services to meet.”**

The Reporting Committee should gather and compile rigorous reports on legal harms from a broad range of service providers, civil society organisations, academic research, ‘super-complaint’ groups and court cases on illegal content. The Reporting Committee would likewise consist of experts and independent members. When appointing the Reporting Committee, the Board should take extra care not to simply transplant multiple members from other parts of Ofcom who have previously operated in areas where they may play a much more active censorship role.

Ofcom’s online harms regulation should be underpinned by principles of international human rights law, using the established UN human rights framework to set out public interest objectives for online services to meet. If these recommendations are followed, Parliament’s comments can feed into the annual review of practices, through which amendments could be made to the work programme to keep pace with changes in technology and user demand. The frequency of this reporting should make sure that any teething problems can be dealt with quickly and with the input of external expertise. In the long term, this will help to make sure that Ofcom is a dynamic body that can adapt to new threats and modes of doing business online.

Functions

Ofcom’s core functions would be to:

- perform proactive, independent research;
- conduct inquiries into emerging problems;
- make recommendations to Parliament;
- impose penalties for serious and repeated failure to implement processes.

The research function of the regulator should facilitate access for smaller players to ‘off the shelf’ technical solutions. Stakeholders should use this function to share information on how to proactively identify and flag accounts displaying suspicious patterns of behaviour. It should also inform what areas platforms feel they need greater clarity over to back up their own terms and conditions for legal content.

Sites should also not face penalties for actively identifying problematic but legal content on their services: indeed, they should be encouraged to dig deeper. This model avoids firms facing liability for tackling their own problems head-on. The European Commission has likewise said that sites taking proactive steps to prevent inappropriate content should not be regarded as assuming liability for it.

Policing

The Conservative Party's 2019 manifesto was confident in its pitch for the UK to embrace new ways of detecting and investigating crime through technology. It stated that:

> 'We will embrace new technologies and crack down on online crimes. We will create a new national cyber crime force and empower the police to safely use new technologies like biometrics and artificial intelligence, along with the use of DNA, within a strict legal framework. We will also create a world-class National Crime Laboratory...'[101]

For most of the serious online harms listed in the scope of the White Paper, the physical harm is happening offline, with details about it being conveyed online. For CSAI, the primary concern is the child physically being abused. For terrorist content, it is the acts of terrorism. For modern slavery, hate crime, violence, sale of illegal goods etc – the most serious element of the harm is usually being perpetrated by people to one another. The online platforms are a means of communicating that crime rather than the act itself.

This paper does not seek to prescribe exactly what policing tools forces should be using to catch criminals online, not least because criminals are constantly adopting new and more sophisticated ways of operating. It is a matter for Police and Crime Commissioners to decide what they need along with the various national taskforces and agencies that have responsibilities for different areas of crime. HM Inspectorate of Constabulary and Fire & Rescue Services conducted an inspection of the police response to cyber-dependant crime.[102] It is envisioned that the category of cyber-enabled crimes would incorporate illegal online harms as they are broadly defined as, 'existing crimes that have been transformed in scale or form by the use of the Internet'. It is not clear whether the 'new national cyber crime force' will sit under the National Crime Agency or within existing regional police forces, but it is clear either way that substantial investment will be needed.

For threats to national security and child sexual exploitation and abuse, the White Paper adds government power to direct these codes and require sign-off from the Home Secretary. This is an unnecessary overreach of government into the independence of the regulator. Police and intelligence agencies have already agreed expedited access to electronic communications sent by terrorists, serious crime gangs and white-collar criminals. Under a new agreement between the UK and the US signed in October 2019,[103] US technology companies (including Facebook, Google and Twitter) will be compelled hand over the content of emails, texts and direct messages to UK law enforcement bodies and vice versa.[104]

[101] The Conservative and Unionist Party, "Manifesto 2019". Available from: https://assets-global.website-files.com/5da42e2cae7ebd3f8bde353c/5dda924905da587992a064ba_Conservative%202019%20Manifesto.pdf.

[102] HM Inspectorate of Constabulary and Fire & Rescue Services, "Cyber: Keep the light on". Available from: https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/cyber-keep-the-light-on-an-inspection-of-the-police-response-to-cyber-dependent-crime.pdf.

[103] The deal is yet to be ratified in either Congress or the House of Commons, but is expected to come into force early this year. Available from: https://www.gov.uk/government/news/uk-and-us-sign-landmark-data-access-agreement.

[104] The UK has obtained assurances which are in line with the government's continued opposition to the death penalty in all circumstances.

Conclusion

Our specialised model of regulation provides a credible alternative to the course set by the White Paper.

By focusing on what indisputably causes harm – the most insidious and illegal acts – and providing the police with the right resources to tackle them, it allows the law to be fairly applied before holding platforms to account. The standards it sets are democratically decided rather than applied on the whim of the regulator, and apply both online and offline. It safeguards against creeping censorship that could seriously affect our right to legally express ourselves online and to do business without being priced out the market by compliance costs.

In addressing the most pressing problems instead of seeking to tackle a wide breadth of legal and illegal content under one system, our model is also likely to be far more effective in delivering on the Government's ambitions in this area – and gaining public trust.

Greater resources can go to the police to seek out those abusing children rather than to bureaucrats in an already bloated regulator. Users can continue to enjoy the voice that the internet gives them without fear of being censored. A new generation of online companies can help our economy grow through the challenging years ahead, unimpeded by red tape and uncertainty about whether their new breed of services may be in scope.

Regulators inevitably grow and adapt to changing patterns in the sector they cover, but it is very hard to remove powers once they have already been vested. Far better to focus on the illegal issues which you can clearly identify than to try to make firms liable for almost anything and undermine the lawfulness of the entire regulator.

Rather than leaving it up to a regulator to decide on what is harmful, covering everything from terrorism to misogyny to excessive screen time, it should be up to the Government to carefully – and with the full scrutiny of Parliament – make new laws that preserve the distinction between legal and illegal, and fit the online and offline worlds alike.

Centre
for Policy
Studies

57 Tufton Street, London, SW1P 3QL
September 2020
ISBN 978-1-910627-90-7