CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

RESPONSE TO COMPETITION AND MARKETS AUTHORITY CALL FOR EVIDENCE ON DIGITAL MERGERS

1. We have worked extensively on online harms issues over the past year and our work has been influential on government and parliament. This short note lists three summary observations that might be of interest to the Authority and describes the work we have done in the adjacent area of online harm reduction, on which we have also attached a fuller paper for reference.

User-generated content tech platforms are products of regulation

2. In our extensive work on the regulation of technology companies, we have found it helpful to strike down the myth that modern digital platform companies are bastions of entrepreneurial competition. The big tech platforms that host content made by others are fundamentally creatures of regulation. Regulation shields them from much of the responsibility of newspapers, TV companies, radio stations etc from what people post on their platforms. The user generated content platforms are dependent on rules in the USA, NAFTA and Europe that make them a 'mere conduit' for the material of others. It has allowed them to grow to colossal size without having to invest much (compared to their revenues) in responsibility. Back in 2000 when less than 5% of the population had used the internet and no one knew what would happen this approach made sense. It no longer does today.

Algorithmic bias has been regulated since 1984

3. It is often forgotten that competition authorities in the USA and the EU/EC have a long history of regulating to mitigate competition and societal harms arising from electronic networks that use algorithms to display information. Regulatory work on algorithms used to anti-competitive advantage in Computerised Reservation Systems for airline ticketing go back to the 1980s. In 1984, Congress introduced rules[1] to combat 'screen bias' in how information was displayed on information systems run by airlines. There were then only a few ticketing systems (essentially two) and no market entry. The EC and then EU introduced extensive regulations thereafter including Regulation No 2299/89 in 1989.

4. This body of work also formed the intellectual basis in the 1990s for UK, then EU policy (Access Directive 2002/19/EC) on EPG regulation for due prominence in which William Perrin now a Trustee of Carnegie UK Trust was involved as a DTI and Downing Street civil servant. The Access Directive draws out the inter-relationships between competition law and wider societal issues that the CMA has noted in its Digital Markets Strategy:

   'Competition rules alone may not be sufficient to ensure cultural diversity and media pluralism in the area of digital television. (Recital 10)'

[1] United Press International (Feb. 9, 1984) 'Government to issue air reservation rules' by Judith Dugan. See https://www.upi.com/Archives/1984/02/09/Government-to-issue-air-reservation-rules/2723445150800/

5. OFCOM went on in 2004 to set rules for EPGs reflecting competition and broadcasting objectives and continues to keep them in force, last updated in 2014[2]. The amendment of the Audio Visual Media Services Directive preserves the ability of Member States to maintain EPGs[3] (Recital 25):

> *"Directive 2010/13/EU is without prejudice to the ability of Member States to impose obligations to ensure the appropriate prominence of content of general interest under defined general interest objectives such as media pluralism, freedom of speech and cultural diversity. Such obligations should only be imposed where they are necessary to meet general interest objectives clearly defined by Member States in accordance with Union law. Where Member States decide to impose rules on appropriate prominence, they should only impose proportionate obligations on undertakings in the interests of legitimate public policy considerations."*

<u>No awareness of price paid by consumer</u>

6. Useful survey work by Doteveryone[4] suggests that people have little conception of price paid. In our paper (see below) we say that:

> "*Indeed, there is a good case to make for market failure in social media and messaging services – at a basic level, people do not comprehend the price they are paying to use a service; research by doteveryone revealed that 70% of people 'don't realise free apps make money from data', and 62% 'don't realise social media make money from data' . Without basic awareness of price and value amongst consumers it will be hard for a market to operate efficiently, if at all, and this market is currently one which sees a number of super-dominant operators*."

This is intrinsic to assessing whether competition can ever work in such an environment. With no awareness of price by the consumer, the service provider can continue to extract surplus far beyond marginal cost with little or no response from the consumer to what I paid for goods would be a strong signal to switch supplier if commensurate consumer value increases were not obtained.

<u>Children as valuable customers – special competition considerations?</u>

7. It is unusual for the CMA to consider a market where children (people under 18 years old) are especially active as consumers. A number of experts have flagged up the special vulnerability of children as consumers of digital products and services and design features of digital platforms apparently designed to manipulate children's behaviour[5]. We would suggest that the CMA pay special attention to the effects of the digital advertising and related markets on children and to consider whether there are special competition issues. For instance the fundamental ability to give meaningful consent as well as market definition – is there a separate market for digital advertising to children, does market concentration have a

---

[2] Ofcom Code on Electronic Programme Guides https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/epg-code

[3] DIRECTIVE (EU) 2018/1808 https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018L1808&from=EN

[4] See https://understanding.doteveryone.org.uk/

[5] See the work of 5Rights Foundation https://5rightsfoundation.com/resources.html and Sonia Livingstone et al at LSE https://blogs.lse.ac.uk/mediapolicyproject/2019/07/09/data-and-privacy-in-the-digital-age-from-evidence-to-policy/ also the UNICEF paper 'Children and Digital Advertising' December 2018

particularly heavy impact on children? Brands have for decades made huge efforts to influence the behaviour of the young and establish lifetime patterns of consumption. Special rules exist for children's advertising in many media forms that reflect their developmental inability to make the same judgements as adults. The advent of digital media has enabled the aggregation of data about young people. The Age Appropriate Design Code now emerging from the ICO applies special rules to data use.

8. We feel that there would be a substantial gap in the CMA's work if it did not give special consideration to the competition effects on children.

Our work on social media harm reduction

9. This note also covers a full reference paper that sets out work we have carried out to develop a proposal for a statutory duty of care for harm reduction on social media.

10. In 2018-2019, Professor Lorna Woods (Professor of Internet Law in the School of Law at the University of Essex) and William Perrin (a Carnegie UK Trustee and former UK government Civil Servant) developed a public policy proposal to improve the safety of some users of internet services in the United Kingdom through a statutory duty of care enforced by a regulator. Woods and Perrin's work under the aegis of Carnegie UK Trust took the form of many blog posts, presentations and seminars.

11. The attached reference paper, drawing together our work on a statutory duty of care was published in April 2019, just prior to the publication of the Online Harms White Paper. It can also be viewed, along with all the other material relating to this proposal and a full recent response to the DCMS consultation on the Online Harms White Paper, on the Carnegie UK Trust website: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/

12. Our work has influenced the recommendations of a number of bodies, including: the House of Commons Science and Technology Committee, the Lords Communications Committee, the NSPCC, the Children's Commissioner, the UK Chief Medical Officers, the APPG on Social Media and Young People and the Labour Party.[6] A statutory duty of care has been adopted – though not fully as we envisaged – by the Government as the basis for its Online Harms White Paper proposals[7]. Most recently, though it did not refer to our work, a report to the French Ministry of Digital Affairs referenced a "duty of care" as the proposed basis for social media regulation.[8]

13. While not directly focused on competition law or the harms that arise from digital mergers, our work has been cognisant of the wider legislative and regulatory context into which any new regulatory model must fit. It addresses the particular challenge posed by new and

---

[6] https://www.nspcc.org.uk/globalassets/ documents/news/taming-the-wild-west-web-regulate-social-networks.pdf; https://www.childrenscommissioner.gov.uk/2019/02/06/childrens-commissioner-publishes-astatutory-duty-of-care-for-online-service-providers/; https://www.gov. uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-ofreviews/; https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm; https://labour.org.uk/press/tom-watson-speech-fixing-distorted-digital-market/; https://www.parliament.uk/business/committees/committees-a-z/lords-select/ communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-toregulate/; https://www.rsph.org.uk/our-work/policy/wellbeing/new-filters.html

[7] https://www.gov.uk/government/consultations/online-harms-white-paper

[8] http://www.iicom.org/images/iic/themes/news/Reports/French-social-media-framework---May-2019.pdf

innovative technologies with reference to the precautionary principle[9], which may be of interest to the CMA in its own deliberations, and also to the established approach of regulating in the public interest for externalities and harms to members of the public. During the consultation period, we have worked closely with other organisations and consumer groups who have an interest in how consumer harms emerge on online platforms (for example, copyright infringement, fake reviews, scams and the sale of unsafe products) and who see the explicit exclusion of economic harms from the DCMS scope as an error. (See our response to the White Paper consultation for more detail on why we think this should be included[10].)

14. Another angle that is relevant to this enquiry is the fact that the desire to gain data for advertising revenue has driven at least some of the problematic design choices of the major platforms; for example, a focus on user engagement as a business priority means that content that gets user engagement is rewarded, which then drives more and more extreme content (on whatever topic the user is engaged in). This then becomes exacerbated by the size of the major platforms: they are sufficiently large that they have difficulty in keeping on top of the problem, and – even where they make headway with a significant proportion of problematic content, the remainder will still be a big issue. Our duty of care proposals have relevance here: it seeks not just to tackle dealing with problems once they've arisen but also to address the conditions that shape the way content is created/shared. Design choices around frictionless communication also influence the ease with which content can spread across platforms.

15. Finally, given that many of the harms we focus on in our work have a societal impact – such as the impact on democracy of the spread of disinformation and the abuse or intimidation of public figures – there may also be a case to be made for an extension of the public interest test found in ss 42 *et seq* Enterprise Act to apply to mergers in this context.

16. There are many moving parts in this landscape, and many government and regulatory organisations undertaking concurrent reviews of bits of it. Protecting users from harm – however it manifests itself - has to be at the heart of all those proposals. The dominance of a small number of platforms is part, but not all, of the problem and a statutory duty of care does not displace the design and implementation of competition law that is fit for the digital age. However, we would urge the CMA to ensure that any new regulatory regime takes account of its principles and seeks to join up with the online harms regulator at the earliest opportunity.

17. We are happy to speak to you further about our proposals or assist in any way in the next phase of the CMA's review.

Carnegie UK Trust

July 2019

[Attachment: "Online Harm Reduction: a statutory duty of care and a regulator" (April 2019)]

---

9 United Kingdom Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA), The Precautionary Principle: Policy and Application, available: http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm

10 https://www.carnegieuktrust.org.uk/publications/response-to-the-online-harms-white-paper/

CarnegieUK TRUST

CHANGING MINDS • CHANGING LIVES

Online harm reduction – a statutory duty of care and regulator

April 2019

About this report

The authors

The authors Lorna Woods and William Perrin have vast experience in regulation and free speech issues.

Lorna is Professor of Internet Law in the School of Law at the University of Essex and a member of the Human Rights Centre there. She started her career as a solicitor focussing on the TMT sectors. On becoming an academic, her research areas have lain in these fields. Recent publications include: 'Video-sharing platforms in the revised Audiovisual Media Services Directive' (2019) 23 Comms Law 127; 'Competition Law and Telecommunications' in Telecommunications Law and Regulation Walden (ed) (5th ed); 'Digital Freedom of Expression' in Douglas-Scott and Hatzis (eds) Research Handbook on EU Law and Human Rights. She currently teaches internet and data protection law and copyright law, but has also taught media law, competition law and EU law. She was a solicitor in private practice specialising in telecoms, media and technology law.

William has worked on technology policy since the 1990s, was a driving force behind the creation of OFCOM and worked on regulatory regimes in gas and electricity, alcohol and entertainment licensing, gambling and many economic and social sectors while working in the UK government's Cabinet Office, Downing Street and Department of Trade and Industry. William is a trustee of Carnegie UK Trust and several other charities active in the digital philanthropy.

The authors are extremely grateful to Carnegie UK Trust for their support in this work, in particular Carnegie Associate Maeve Walsh and Carnegie executives Douglas White and Anna Grant for their support with this project.

Carnegie UK Trust

Carnegie UK Trust was established in 1913 by Scottish-American industrialist and philanthropist Andrew Carnegie to seek:

"Improvement of the well-being of the masses of the people of Great Britain and Ireland by such means as are embraced within the meaning of the word "charitable" and which the Trustees may from time to time select as best fitted from age to age for securing these purposes, remembering that new needs are constantly arising as the masses advance."

The authors worked on this report pro-bono; Carnegie UK Trust eventually bought out some of Lorna Woods' time from the University of Essex.

This report

In 2018-2019 Professor Lorna Woods and William Perrin developed a public policy proposal to improve the safety of some users of internet services in the United Kingdom through a statutory duty of care enforced by a regulator. Woods and Perrin's work under the aegis of Carnegie UK Trust took the form of many blog posts, presentations and seminars.

This paper consolidates that work into a single document.

This report is being published in advance of the publication of the UK Government's Online Harms White Paper, which is expected to set out comprehensive proposals for regulation in this area.

Table of Contents

Executive Summary

The main aspects of our proposal are set out in summary below. Supporting detail follows in the subsequent chapters.

We have designed a new regulatory regime to reduce harm to people from social media services. We draw on the experience of regulating in the public interest for externalities and harms to the public in sectors as diverse as financial services and health and safety. Our approach – looking at the design of the service - is systemic rather than content-based, preventative rather than palliative.
At the heart of the new regime would be a 'duty of care' set out by Parliament in statute. This statutory duty of care would require most companies that provide social media or online messaging services used in the UK to protect people in the UK from reasonably foreseeable harms that might arise from use of those services. This approach is risk-based and outcomes-focused. A regulator would have sufficient powers to ensure that companies delivered on their statutory duty of care.

Duty of care

Social media and messaging service providers should each be seen as responsible for a public space they have created, much as property owners or operators are in the physical world. Everything that happens on a social media or messaging service is a result of corporate decisions: about the terms of service, the software deployed and the resources put into enforcing the terms of service and maintaining the software. These design choices are not neutral: they may encourage or discourage certain behaviours by the users of the service.

In the physical world, Parliament has long imposed statutory duties of care upon property owners or occupiers in respect of people using their places, as well as on employers in respect of their employees. A statutory duty of care is simple, broadly based and largely future-proof. For instance, as we set out in chapter 5, the duties of care in the Health and Safety at Work Act 1974[1] still work well today, enforced and with their application kept up to date by a competent regulator. A statutory duty of care focuses on the objective – harm reduction – and leaves the detail of the means to those best placed to come up with solutions in context: the companies who are subject to the duty of care. A statutory duty of care returns the cost of harms to those responsible for them, an application of the micro-economically efficient 'polluter pays[2]' principle. The E-Commerce Directive[3], permits duties of care introduced by Member States; the Audiovisual Media Services Directive (as amended in 2018) requires Member States to take some form of regulatory action in relation to a sub-set of social media platforms – video-sharing platforms.

The continual evolution of online services, where software is updated almost continuously makes traditional evidence gathering such as long-term randomised control trials problematic. New services. adopted rapidly that potentially cause harm illustrate long standing tensions between science and public policy. For decades scientists and politicians have wrestled with commercial actions for which there is emergent evidence of harms: genetically modified foods, human fertilisation and embryology, mammalian cloning, nanotechnologies, mobile phone electromagnetic radiation, pesticides, bovine spongiform encephalopathy. In 2002, risk management specialists reached a balanced definition of the precautionary principle that allows economic development to proceed at risk in areas where there is emergent evidence of harms but scientific certainty is lacking within the time frame for decision making.

Emergent evidence of harm caused by online services poses many questions: whether bullying of children is widespread or whether such behaviour harms the victim[4]; whether rape and death threats to women in public life has any real impact on them, or society; or whether the use of devices with screens in itself causes problems. The precautionary principle[5], which we describe in chapter 2, provides the basis for policymaking in this field, where evidence of harm may be evident, but not conclusive of causation. Companies should embrace the precautionary principle as it protects them from requirements to ban particular types of content or speakers by politicians who may over-react in the face of moral panic.

Parliament should guide the regulator with a non-exclusive list of harms for it to focus upon. We set these out in chapter 6: the stirring up offences including misogyny, harassment, economic harm, emotional harm, harms to national security, to the judicial process and to democracy. Parliament has created regulators before that have had few problems in arbitrating complex social issues; these harms should not be beyond the capacity of a competent and independent regulator. Some companies would welcome the guidance.

Who would be regulated?

Chapter 7 sets out our thinking on the services that would fall under a statutory duty of care. The harms set out by the UK government in its 2017 Green Paper[6] and by civil society focus on social media and messaging services, rather than basic web publishing or e-commerce.. We propose regulating services that:

- Have a strong two-way or multiway communications component;
- Display user-generated content publicly or to a large member/user audience or group.

We exclude from the scope of our proposals those online services already subject to a comprehensive regulatory regime – notably the press and broadcast media. But include messaging services that have evolved and permit large group sizes and semi or wholly public groups.

The regime would cover reasonably foreseeable harm that occurs to people who are users of a service and reasonably foreseeable harm to people who are not users of a service.

Our proposals do not displace existing laws whether regulatory in nature – for example, the regulatory regime overseen by the Advertising Standards Authority, civil (e.g. defamation), or criminal (e.g. gaslighting and stalking).

Duty of care – risk-based regulation

Central to the duty of care is the idea of risk, as discussed in chapter 9. If a service provider targets or is used by a vulnerable group of users (e.g. children), the risk of harm is greater and the service provider should have more safeguard mechanisms in place than a service which is, for example, aimed at adults and has community rules agreed by the users themselves (not imposed as part of ToS by the provider) to allow robust or even aggressive communications. This is a systemic approach, not a palliative one that merely attempts to clean up after the problem.

Regulation in the UK has traditionally been proportionate, mindful of the size of the company concerned and the risk its activities present. Small, low-risk companies should not face an undue burden from the proposed regulation. Baking in harm reduction to the design of services from the outset reduces uncertainty and minimises costs later in a company's growth.

How would regulation work?

Chapters 9 and 10 describe how the regulatory regime would work. We argue that the regulator should be one of the existing regulators with a strong track record of dealing with global companies. We favour OFCOM as the regulator, to avoid delays in establishing the regime, funded on a polluter pays basis by those it regulates.

The regulator should be given substantial freedom in its approach to remain relevant and flexible over time. We suggest the regulator employ a harm reduction method similar to that used for reducing pollution: agree tests for harm, run the tests, the company responsible for harm invests to reduce the tested level, test again to see if investment has worked and repeat if necessary. If the level of harm does not fall or if a company does not co-operate then the regulator will have sanctions. We know that harms can be surveyed and measured as shown by OFCOM and the ICO in their survey work of September 2018[7].

In a model process, the regulator would work with civil society, users, victims and the companies to determine the tests and discuss both companies harm reduction plans and their outcomes. The regulator would have a range of powers to obtain information from regulated companies as well as having its own research function. The regulator is there to tackle systemic issues in companies and, in this proposal, individuals would not have a right to complain to the regulator or a right of action to the courts for breach of the statutory duty of care. We suggest that a form of 'super-complaint' mechanism, which empowers designated bodies to raise a complaint that there has been a breach of statutory duty, be introduced.

What are the sanctions and penalties?

The regulator needs effective powers to make companies change behaviour. We propose large fines set as a proportion of turnover, along the lines of the General Data Protection Regulation (GDPR)[8] and Competition Act[9] regimes. Strong criminal sanctions have to be treated cautiously in a free speech context. We have also made suggestions of powers that bite on directors personally, such as fines.

Timing

We conclude, in chapter 11, that urgent action is needed, especially in a fast-changing sector, creating a tension with a traditional deliberative process of forming legislation and then regulation. We would urge the government to find a route to act quickly and bring a statutory duty of care to bear on the companies as fast as possible. There is a risk that if we wait three or four years the harms may be out of control. This isn't good for society, nor the companies concerned.

1. Introduction

This report describes at a high level what a model regulatory regime for harm reduction in social media that respects freedom of expression in the European tradition might look like. The regime we propose is not a silver bullet – it will not solve all internet-related problems. As we shall explain, however, a statutory duty of care enforced by a regulator can provide a solid foundation for more detailed rules or interventions to target specific circumstances; it does not exclude the possibility of other measures (e.g. media literacy and ethics training) being used as well.

The authors began this work in early 2018 after a growing realisation that the many reports of emerging harms from social media had not generated a commensurate response from the regulatory policy community and existing laws (whether civil or criminal) were proving ineffective in countering these harms. We had worked together with Anna Turley MP in 2016 on her Private Members Bill 'Malicious Communications (Social Media)'[10], motivated by the evident impact abuse from social media was having on women in public life. The Committee on Standards in Public Life reported that:

> *'A significant proportion of candidates at the 2017 general election experienced harassment, abuse and intimidation. There has been persistent, vile and shocking abuse, threatened violence including sexual violence, and damage to property. It is clear that much of this behaviour is targeted at certain groups. The widespread use of social media platforms is the most significant factor driving the behaviour we are seeing. Intimidatory behaviour is already affecting the way in which MPs are relating to their constituents, has put off candidates who want to serve their communities from standing for public offices, and threatens to damage the vibrancy and diversity of our public life. However, the Committee believes that our political culture can be protected from further damage if action is taken now.'*[11]

We were delighted when in 2018 Carnegie UK Trust agreed to publish our work, initially as a series of blog posts.

The issues are complex: social media companies have a duty to shareholder interests; individuals do not bear the costs of their actions when they use social media leading to strong externalities; rights to freedom of speech are highly valued but may conflict with each other as well as other fundamental human rights; the issues cut across different global approaches to regulation, making jurisdiction sometimes unclear; and few governments have seemed willing to take a strong lead.

The Government's Internet Safety Strategy Green Paper in Autumn 2017 detailed extensive harms with costs to society and individuals resulting from people's consumption of social media. Companies' early stage growth models and service design decisions appeared to have been predicated on such costs being external to their own production decision. Effective regulation would internalise these costs for the largest operators and lead to a more efficient outcomes for society. Indeed, there is a good case to make for market failure in social media and messaging services – at a basic level, people do not comprehend the price they are paying to use a service; research by doteveryone revealed that 70% of people 'don't realise free apps make money from data', and 62% 'don't realise social media make money from data'[12]. Without basic awareness of price and value amongst consumers it will be hard for a market to operate efficiently, if at all, and this market is currently one which sees a number of super-dominant operators.

UK journey to regulation

We set out below a timeline of the key events from the publication of the UK Digital Economy Bill in 2016 to the present, where we await the publication of the UK Government's Online Harms White Paper.

July 2016	Digital Economy Bill[13] proposes measures to protect children, such as age verification for pornography
Late 2016	Anna Turley MP puts forward "Malicious Communications (Social Media) Bill" [14], with support from the authors.
April 2017	Digital Economy Act 2017 receives Royal Assent.
May 2017	Conservative Party Election Manifesto[15] sets out a proposal for a range of measures to combat harms through a code of practice for social media, clearly setting out a political judgement that there were a wide range of harms arising from online media from which people required protection.
July 2017	Conservatives, as the largest party in Parliament, form an administration.
October 2017	UK Department for Digital, Culture, Media and Sport (DCMS) publish their Internet Safety Strategy Green Paper[16] for consultation. The Green Paper sets out a list of harms rather than definitive policy options to address them.
January 2018	Following discussions with charities and advocacy groups in late 2017, the authors approach Carnegie UK Trust with a proposal to write pro bono a series of exploratory blog posts considering the regulatory options for the Government to reduce social media harms.
January 2018	The Prime Minister, Theresa May, signals in her speech at Davos that the government was serious in pursuing the manifesto commitments.[17]
February 2018	BBFC designated as age verification regulator (for pornography) under Digital Economy Act.[18]
Early 2018	The authors publish a preliminary proposal for a statutory duty of care in a series of posts on the Carnegie UK Trust website. [19]
March 2018	House of Lords Communications Select Committee launches inquiry on Internet Regulation[20]
May 2018	DCMS publishes its response to the consultation on the Internet Safety Strategy Green Paper [21] and signals its intention to bring forward proposals for new laws to tackle a wide range of internet harms.
May 2018	Data Protection Act 2018[22], which includes the implementation of the General Data Protection Regulation, also sets out provisions for an Age-Appropriate Design Code.
June 2018	Information Commissioner's Office (ICO) consults on the Age-Appropriate Design Code.[23]
June-Dec2018	The authors continue to work on the proposal pro bono for Carnegie UK Trust, supported by Carnegie Associate Maeve Walsh from autumn 2018, meeting with a wide range of stakeholders (social media platforms, advocacy groups, politicians, regulators) and present evidence to various consultations and Parliamentary inquiries.
October 2018	BBFC publishes Guidance on Age-verification Arrangements under Digital Economy Act
December 2018	Statutory instruments relating to the meaning of 'commercial basis' (which defines the scope of the age verification rules) as well as guidance from the BBFC on age verification (s. 14 DEA) and 'ancillary services' were laid before Parliament.[24] Debates

	make clear that although pornography may be available on social media platforms (e.g. Twitter) they do not fall within the scope of the DEA regime.[25]
January 2019	Substantial revision of the statutory duty of care published[26], incorporating feedback from stakeholder engagement and new thinking. The Online Pornography (Commercial Basis) Regulations 2019 (SI 2019/23) under s. 14 DEA made but not yet in force.
January-March 2019	Support for a statutory duty of care builds, with endorsements from: the House of Commons Science and Technology Committee, the Lords Communications Committee, the NSPCC, the Children's Commissioner, the UK Chief Medical Officers, the APPG on Social Media and Young People and the Labour Party.[27]
March 2019	Original and new thinking on the Carnegie UK Trust proposal brought together in this published report. UK Government's White Paper, along with details of the new Age Appropriate Design Code, expected imminently.

2. The precautionary principle

One of the recurrent arguments put forward for not regulating social media and other online companies is that they are unique or special: a complex, fast-moving area where traditional regulatory approaches will be blunt instruments that stifle innovation and require platform operators to take on the role of police and/or censors. Another is that the technology is so new, sufficient evidence has not yet been gathered to provide a reliable foundation for legislation; where there is a body of evidence of harm, in most cases the best it can do is prove a correlation between social media use and the identified harm, but not causation.[28]

We believe that the traditional approach of not regulating innovative technologies needs to be balanced with acting where there is good evidence of harm. The precautionary principle provides a framework for potentially hazardous commercial activity to proceed relatively safely and acts as a bulwark against short term political attempts to ban things in the face of moral panic.

Rapidly-propagating social media and messaging services, subject to waves of fashion amongst young people in particular, are an especial challenge for legislators and regulators. The harms are multiple, and may be context- or platform- specific, while the speed of their proliferation makes it difficult for policymakers to amass the usual standard of long-term objective evidence to support the case for regulatory interventions. The software that drives social media and messaging services is updated frequently, often more than once a day. Facebook for instance runs a 'quasi-continuous [software] release cycle'[29] to its webservers. The vast majority of changes are invisible to most users. Tweaks to the software that companies use to decide which content to present to users may not be discernible. Features visible to users change regularly. External researchers cannot access sufficient information about the user experience on a service to perform long term research on service use and harm. Evidencing harm in this unstable and opaque environment is challenging, traditional long-term randomised control trials to observe the effect of aspects of the service on users or others are nearly impossible without deep co-operation from a service provider.

Nonetheless there is substantial indicative evidence of harm both from advocacy groups and more disinterested parties. OFCOM and ICO demonstrated[30] that a basic survey approach can give high level indications of harm as understood by users. So, how do regulation and economic activity proceed in the

face of indicative harm but where scientific certainty cannot be achieved in the time frame available for decision making?

This is not the first time the government has been called to act robustly on possible threats to public health before scientific certainty has been reached. After the many public health and science controversies of the 1990s, the UK government's Interdepartmental Liaison Group on Risk Assessment (ILGRA) published a fully worked-up version of the precautionary principle for UK decision makers.[31]

> *'The precautionary principle should be applied when, on the basis of the best scientific advice available in the time-frame for decision-making: there is good reason to believe that harmful effects may occur to human, animal or plant health, or to the environment; and the level of scientific uncertainty about the consequences or likelihoods is such that risk cannot be assessed with sufficient confidence to inform decision-making.'*

The ILGRA document advises regulators on how to act when early evidence of harm to the public is apparent, but before unequivocal scientific advice has had time to emerge, with a particular focus on novel harms. ILGRA's work focuses on allowing economic activity that might be harmful to proceed 'at risk', rather than a more simplistic, but often short-term politically attractive approach of prohibition. The ILGRA's work is still current and hosted by the Health and Safety Executive (HSE), underpinning risk-based regulation of the sort we propose.

We believe that – by looking at the evidence in relation to screen use, internet use generally and social media use in particular – there is in relation to social media "good reason to believe that harmful effects may occur to human[s]" despite the uncertainties surrounding causation and risk. On this basis we propose that it is appropriate if not necessary to regulate and the following sets out our proposed approach.

3. Overall approach to regulation: system not content

The UK government's Internet Safety Strategy Green Paper set out some of the harms to individuals and society caused by social media users but did not elaborate on concrete proposals to address these problems. Many commentators have sought an analogy for social media services as a guide for the best route to regulation. Whenever social media regulation has been proposed in recent times, it is inevitably discussed in a way that – by framing the question as one of holding the social media platforms accountable for the content on their sites – leads straight into the "platform or publisher" debate: this can be seen in the report of the Committee on Standards in Public Life[32]. This approach is understandable given that the platforms transmit and/or publish messages. It is the approach that existing regimes have adopted; many legal regimes provided for the immunity from legal action of those who are deemed to be mere intermediaries as opposed to publishers.

This debate is not, however, fruitful – it leads to discussions about the removal of immunity and conversely the difficulties of applying content-based regulation in the context of social media given its scale. We agree that there is a difficulty with highly detailed regimes, but in our view the starting point is wrong. The models of publisher and intermediary are an ill-fit for current practice, as recognised by the Report of the DCMS Select Committee into disinformation and fake news[33] – particularly given the need to deploy regimes and enforcement at scale. A new approach is needed. In our view the main analogy for social media networks lies outside the digital realm. When considering harm reduction, social media networks

should be seen as a public or (given that they are privately owned) a quasi-public place – like an office, bar, or theme park. Hundreds of millions of people go to social media networks owned by companies to do a vast range of different things. In our view, they should be protected from harm when they do so.

The analogy of space also works well if we consider the nature of the on-line environment. In this aspect, we have revisited Lessig's work[34]. Lessig observed that computer code sets the conditions on which the internet (and all computers) is used. While there are other constraints on behaviour (law, market, social norms), code is the architecture of cyberspace and affects what people do online: code permits, facilitates and sometimes prohibits. Of course, there are limits to the extent of such control. It is not our argument that we are 'pathetic dots'[35] in the face of engineered determinism. Nonetheless, it is becoming increasingly apparent that the architecture of the platform – as formed by code - also nudges[36] us towards certain behaviour[37], whether this is the intention of the software designer or not. Indeed, there is a concern that when a developer is focussed on a particular objective, he may overlook other interests and possible side-effects[38]. While Lessig's work was oriented along a different line, it reminds us that the environment within which harm occurs is defined by code that the service providers have actively chosen to deploy, their terms of service or contract with the user and the resources service providers deploy to enforce that. While technological tools can be used for positive reasons as well as have negative impacts, it is important to remember that they are not neutral[39], nor are they immutable. Corporate decisions drive what content is displayed to a user. Service providers could choose not to deploy risky services without safeguards or they could develop effective tools to influence risk of harm if they choose to deploy them. The current difficulty seems in part as result of the fact that possible adverse consequences are not considered or, if they are, they are not highly valued. Sean Parker a co-founder of Facebook said in a 2017 interview:

> *'God only knows what it's doing to our children's brains. The thought process that went into building these applications, Facebook being the first of them, ... was all about: How do we consume as much of your time and conscious attention as possible?.... It's a social-validation feedback loop ... exactly the kind of thing that a hacker like myself would come up with, because you're exploiting a vulnerability in human psychology."*[40]

Part of the ethos underpinning the 'by design' approach is the concern to ensure that other issues are taken into account in design and operational choices.[41]

In sum, online environments reflect choices made by the people who create and manage them; those who make choices should be responsible for the reasonably foreseeable risks of those choices.

Our approach, in moving beyond a focus on platform accountability for every item of content on their respective sites, to their responsibility for the systems that they have designed for public use, allows the reduction of harm to be considered within a regulatory approach that scales but, by not focussing directly on types of content, is also committed to the preservation of free speech of all.

In the regime we set out here, oversight would be at a system or platform level, not regulation of specific content – this is significant in terms of placing responsibility on the actions and activities that platform operators control and also in terms of practicality. Regulation at the system level focuses on the architecture of the platform. This is similar to the 'by design' approach seen in data protection and information security (for example in the EU GDPR). We consider this further in the next chapter. It should

be noted that, although a 'by design' approach requires choices at the early developmental stage, it does not mean that questions about the design, deployment and operation of the service can then be forgotten. Ongoing review is important to ensure that the system continues to function as the market and technology develops.[42] The statutory duty of care approach is not a one-off action but an ongoing, flexible and future-proofed responsibility that can be applied effectively to fast-moving technologies and rapidly emerging new services.

In broad terms, our system-level approach looks like this: the regulator would have powers to inspect and survey the networks to ensure that the platform operators had adequate, enforced policies in place to identify risks and to rectify or minimise the risks (though this would not involve constant, general monitoring of platform use). The regulator, in consultation with industry, civil society and network users would set out a model process for identifying and measuring harms in a transparent, consultative way. The regulator would then work with the largest companies to ensure that they had measured harm effectively and published harm reduction strategies addressing the risks of harm identified and mitigating risks that have materialised. Specifying the high-level objectives to safeguard the general public allows room for service providers to act by taking account of the type of service they offer, the risks it poses (particularly to vulnerable users) and the tools and technologies available at the time. The approach builds on the knowledge base of the sector and allows for future proofing. The steps taken are not prescribed but can change depending on their effectiveness and on developments in technologies and their uses.

4 Overarching Legal Frameworks

There are a number of overarching legal frameworks operating at the European level which any regulatory system should take into account, most notably the European Convention on Human Rights (ECHR). Given the uncertainties surrounding Brexit, there are some aspects of the EU legal system that should also be considered: the e-Commerce Directive[43] and the Audiovisual Media Services Directive (as most recently amended).[44] The EU also has a Charter of fundamental rights; the issues raised there are treated as being more or less the same as those under the ECHR and the discussion on the EU Charter is therefore subsumed under the ECHR discussion.[45]

European Convention on Human Rights

The human rights issues surrounding internet use in general and social media more specifically are complex, involving a range of rights held by many different actors whose interests may not be the same. For the purposes of this report, however, we focus only on the central issue: to what extent are regulatory regimes constrained by human rights?

While the impact of the right to freedom of expression is important, it is not the only right that might be relevant. The right to private life is also affected by the use of the internet and social media platforms. Other rights might be implicated too: for example, the right to a fair trial. It should also be noted that there is no hierarchy between the rights – all are equal. In cases of a conflict between rights, an assessment of how an appropriate balance should be made is determined in each instance. In particular, although it will necessarily be implicated in each instance, freedom of expression does not have automatic priority.

The Council of Ministers of the Council of Europe made a declaration[46] on the manipulative capabilities of algorithmic processes in February 2019:

> *'Moreover, data-driven technologies and systems are designed to continuously achieve optimum solutions within the given parameters specified by their developers. When operating at scale, such optimisation processes inevitably prioritise certain values over others, thereby shaping the contexts and environments in which individuals, users and non-users alike, process information and make their decisions. This reconfiguration of environments may be beneficial for some individuals and groups while detrimental to others, which raises serious questions about the resulting distributional outcomes…. The Council of Ministers encourages member States to assume their responsibility to address this threat by [inter alia] considering the need for additional protective frameworks related to data that go beyond current notions of personal data protection and privacy and address the significant impacts of the targeted use of data on societies and on the exercise of human rights more broadly…. taking appropriate and proportionate measures to ensure that effective legal guarantees are in place against such forms of illegitimate interference'*

Whilst not legally, binding such declarations are sometimes used as a guide to interpretation by the European Court of Human Rights.

Freedom of Expression

Freedom of expression is found in Article 10 ECHR (and Article 11 EU Charter). Article 10 comprises two paragraphs – the first specifies the right, the second the exceptions to the right. The right protected is broad: it includes the right to hold opinions as well as to express them; the right to be silent as well as to speak; and the right to receive information. Note also that the last sentence of paragraph also recognises that States may license "broadcasting, television and cinema enterprises".

In terms of scope, Article 10 is broad both as to what constitutes expression and who may claim the right. The Court has repeatedly emphasised that Article 10 applies not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb. Such are the demands of pluralism, tolerance and broadmindedness, without which there is no 'democratic society'.[47] The Court has protected insults (in the context of politicians).[48] In general, the 'limits of acceptable criticism are … wider as regards a politician as such than as regards a private individual',[49] and a similar approach has been taken to large corporations.[50] Article 10 would cover communications mediated by electronic communications networks; recent case law might suggest that the platforms too have a right to freedom of expression (though the question has not been directly addressed by the Court).[51] The notion of a restriction on the exercise of the right is also wide. Any regime would therefore have to satisfy the requirements of Article 10(2) ECHR.

Article 10(2) is understood as requiring that any restriction or interference with the right must be justified according to a three-stage test in which the conditions are cumulative. The conditions are:

- the interference is prescribed by law
- the interference is aimed at protecting one or more of a series of interests listed in Article 10(2):
  - national security;
  - territorial integrity;
  - public safety;

- prevention of disorder or crime;
- protection of health, morals, reputation or rights of others;
- preventing the disclosure of information received in confidence; and
- maintaining the authority and impartiality of the judiciary;

- the interference is necessary in a democratic society – this, in essence, is a form of proportionality test but where 'necessary' has been understood to mean 'pressing social need'.[52]

These conditions have been interpreted strictly and the burden to show that they have been satisfied in a given case lies with the State. Nonetheless, States retain a 'margin of appreciation' to assess whether a need exists and how to tackle it. The margin of appreciation varies depending on the type of speech in issue – the State has wider room for manoeuvre in relation to morals and commercial speech for example than political speech – private speech has not been much discussed in the case law. A wider margin of appreciation also exists when the State is trying to reach a balance between conflicting rights. The outcome of any proportionality assessment will also depend on the objective to be protected, as well as the severity of the interference. The Court has repeatedly held that, even where the speech could otherwise be legitimately sanctioned, excessive sanctions may lead to a breach of freedom of expression[53]; there is a concern about the chilling effect of such sanctions. The criminalisation of speech is a more severe intervention than, for example, the use of civil damages. Prior censorship gives rise to grave concerns, though injunctions may be permissible where the information is not already available[54]. In Yildirim, a court ordered a blocking order in respect of certain hosting sites because some of the content on those sites was illegal.[55] The order was challenged successfully. The order prevented third parties, who were not breaking the law, from accessing their own materials. This the Court described as 'collateral censorship' and it rendered the State's action disproportionate.

One exception to the wide protection of Article 10 is expression which falls outside Article 10 by virtue of Article 17 ECHR which prevents Convention rights from being relied on to undermine values underpinning the Convention[56]. Article 17 has typically been used in the context of anti-semitism and other forms of hate speech. The effect of Article 17 is that a State does not need to justify its restrictions (as it would under Article 10(2)). For example, in *Belkacem v Belgium*[57], a user of YouTube was prosecuted for making grossly inflammatory statements about the then Minister of Defence of Belgium and that he had harassed the husband of a Belgian politician after her death by posting a video saying that she would spend eternity in Hell. The claimant argued that he was just expressing himself in a shocking way; the Court however took the view that he had sought to incite hatred, discrimination and violence against non-Muslims[58] and that his speech lay outside the protection of Article 10 altogether by virtue of Article 17 ECHR. The claim was inadmissible. Article 17 is consequently seen by some as problematic[59] - it should not be interpreted expansively.

Article 16 ECHR expressly authorises restrictions on the political activities of aliens even though those restrictions may interfere with freedom of expression under Article 10 as well as other Convention freedoms.

Article 10, private actors and positive obligations

Article 10 classically applies to actions of the State. It is a negative obligation; that is, it operates to stop States from imposing restrictions; sometimes the actions of private actors may be attributable

to the State. States may also be under positive obligations to protect rights, including taking steps in relation to the actions of private actors. In *López-Ostra v. Spain*, the Court held that a positive obligation requires public authorities to take 'reasonable and appropriate measures to secure' the rights of an individual.[60] One question which arises, but which has not been addressed by the Court, is the extent to which hate speech – and perhaps other forms of aggressive speech – has a silencing effect, particularly in relation to minorities and vulnerable groups[61]. In her thematic report on online and Information and Communications Technology (ICT) facilitated violence against women from a human rights perspective[62], Dubravka Šimonović, the UN Special Rapporteur on Violence against Women, noted that women in public life – whether this be parliamentarians, journalists and the like - are particularly targeted by online and ICT-facilitated violence. Note that the Court has held, albeit in the context of the regulation of broadcasting, that the State is the ultimate guarantor of pluralism; it may therefore be obliged to act in the interests of pluralism,[63] in addition to acting in protecting the rights of others.

There is another question: that of whether a platform, in imposing its own standards[64], violates its users' freedom of expression.[65] This would seem to come down to the question of whether the user has a right to the platform (over and above that derived from contract).[66] The most that can be inferred from existing case law seems to be a right not to be discriminated against in terms of the media company's decision as to whether to accept the content or not. In *Tierfabriken*, which concerned the refusal to broadcast an advertisement regarding animal rights, the Court found that there was no such justification for refusal, even though there were other tv channels on which the claimant could broadcast the advert. Of course, for issues which are sensitive, there may be reasons to justify such a refusal.[67] As regards the press, the European Court of Human Rights held that, 'the State's obligation to ensure the individual's freedom of expression does not give private citizens or organisations an unfettered right of access to the media in order to put forward opinions'.[68] While the Court accepts that in some instances the State may be under an obligation to ensure access to privately owned spaces so as to facilitate speech, this is not an automatic right that triumphs over those of the property-owner. The case of *Appleby* concerned the rights of protesters who were denied access to a privately-owned shopping centre where they wished to hand out leaflets. The ECtHR held that there had been no violation of Article 10. Factors in deciding whether a positive obligation under Article 10 exists include the kind of expression rights at stake; their capability to contribute to public debates; the nature and scope of restrictions on expression rights; the ability of alternative venues for expression; and the weight of countervailing rights of others or the public.[69] The claimants in *Appleby* lost, even though they were engaging in speech relating to a matter of public interest. One factor was the existence of other venues.

In the context of social media, there are a range of platforms to choose from and much of the speech is not directed at matters of great moment; it is unclear the extent to which any claim to a particular platform would be successful. Indeed, while the Court has recognised the importance of the internet and referred to the right to internet access in relation to State censorship[70], the position as regards access to the internet as against private individuals is still uncertain; it may be that rights of access might arise only where the claimants were particularly vulnerable.[71]

Article 8 ECHR

Article 8 protects individuals' privacy, but the provision's scope is far broader than confidentiality. Indeed, Article 8 does not refer to privacy but to a person's 'private and family life, his home and his correspondence'; it also implicitly includes data protection concerns[72]. There is no exhaustive definition but the Court has held that:

> *'"private life" extends to aspects relating to personal identity, such as a person's name or picture, and furthermore includes a person's physical and psychological integrity … [and] a zone of interaction with others, even in a public context …'.*[73]

Against this broad understanding, Article 8 is not just about the accumulation of data and the profiling of individuals, but may also include individuals' right to maintain familial relationships and friendships. (and may be a better fit for 'private speech' than Article 10).[74] As with Article 10, Article 8 is not unlimited, but any State interference must be justified by reference to the same three stage test we saw in relation to Article 10(2).

In addition to concerns about defamation and the misuse of private information, Article 8 may be relevant in terms of some of the abusive communications. As with Article 10, Article 8 imposes positive obligations on States and States may therefore be required to act to protect individuals from psychological damage.[75] A key case is KU.[76] KU was a child. Someone placed an advertisement on an internet dating site in his name without his knowledge. The advertisement described him, as well as giving a link to a web page which showed his picture. The advertisement stated that he was looking for an intimate relationship with a boy of his age or older. An older man approached him as a result. The Finnish authorities could not trace the perpetrator due to rules protecting anonymous speech, although the service provider could have been penalised under national law. The ECtHR confirmed that Article 8 was applicable, insofar as the case concerned KU's moral integrity and there was a severe threat. Here, effective deterrence in the view of the ECtHR required 'efficient criminal-law provisions'. The case law indicates that physical damage is not required to trigger the State's obligations, though whether additional factors – notably the vulnerability of the victim – is necessary to trigger them is less clear – to date the case law has mainly been concerned with minorities and vulnerable individuals (those with mental handicaps and children).[77]

Application to a Statutory Duty of Care

Assessing the proposal in terms of Article 10, it has the potential to constitute an interference with speech. We should, however, remember that the right to freedom of expression is not absolute and the State may (in some instances must) take action to protect other interests/rights. The right to private life is an obvious example here, but also security (which might justify actions against terrorism) and the right to life, as well as more generally the prevention of crime (for example, child pornography). It is highly likely that a regulatory scheme for the purposes of ensuring online safety would therefore be in the public interest as understood in relation to Article 10(2). The question would then be about the proportionality of the measure, as well as its implementation. In terms of the proposal, no particular content is targeted by the regulation; indeed, no types of speech are so targeted– in that sense it is content neutral. Indeed, the UN Rapporteur on Freedom of Expression, David Kaye, and the UN Special Rapporteur on Violence against Women note that "steps to tackle the abuse-enabling environments often faced by women online" [78] which suggests a systems approach could be part of a human rights respecting mechanism for tackling such harms. Insofar as a range of mechanisms are envisaged whereby a social media platform could satisfy the duty of care that do not involve take down or filtering of the content by the platform (though users should be free not to see content if they choose not to) – for example, age verification, reporting mechanisms the impact on speech is likely to be proportionate even though speech that is less extreme than hate speech may be affected. Furthermore, the proposal envisages that there will be a range of platforms (as there currently are) and that these will have different standards and approaches, so allowing space for different ways of communicating. While the possibility remains that a recalcitrant, refusenik company may find itself receiving high penalties (possibly even criminal penalties), the aim of the

proposal is to facilitate best practice and only to take regulatory enforcement action as a last resort, again emphasising the proportionality of the structure. This, of course, does not suggest that rights should be treated cavalierly but simply that the rights framework does not seem to rule out such regulatory action. We return to some particular issues in this respect below.

Other Human Rights Frameworks

The United Nations Convention on Rights of a Child requires a balance to be struck between a child's right to free speech and their right to be protected from harmful content. The Convention says that signatories shall:

> *'Encourage the development of appropriate guidelines for the protection of the child from information and material injurious to his or her well-being,'*[79]

E-Commerce Directive

The e-Commerce Directive sets out to provide a coherent framework in which Member States could regulate "information society services" (ISS), but in which the free flow of those services through the internal market would be facilitated – specifically where differences in regulatory approach between Member States might constitute a barrier to trade. While a horizontal measure, some fields lie outside its scope: taxation, data protection and gambling. Furthermore, telecommunications and broadcasting did not constitute ISS. In addition to identifying which State would have the responsibility for regulating which services (Article 3), the directive provided that: the taking up and pursuit of the activity of providing ISS should not be subject to prior authorisation;[80] electronic contracting should be allowed; certain minimum information must be required to be provided by the ISS providers; and that immunity for certain categories of ISS providers on specified conditions should be provided. The fact that the directive is a minimum harmonisation directive means that Member States may impose more stringent rules on operators established within their respective territories, though these must be notified to the Commission. Operators established elsewhere could not be subject to such obligations save in the instance that a derogation might apply.

Intermediary Immunity from Liability

The e-Commerce Directive does not harmonise liability – as noted, this is for individual Member States to decide – but rather the limitations on any such liability for matters falling within the scope of the directive. The aim of taking action was to harmonise the level of liability applying across the EU, and to avoid barriers to trade as well as providing reassurance about levels of liability.[81] So beyond some basic requirements, for example, information to be provider by ISS providers, the e-Commerce Directive is silent as to standards of care to be attained by those providers (unless the ISS provider forms part of a regulated profession).

The provisions harmonising limitations on liability are found in Articles 12-15 e-Commerce Directive. Although these provisions are often referred to as provisions related to intermediaries, there is no definition of an intermediary in the e-Commerce Directive. Rather the directive envisages that ISS providers which carry out certain functions – as briefly described in the relevant provision – will receive immunity. The e-Commerce regime this divides intermediary services into three categories:

- 'mere conduit' (Article 12);
- 'caching' (Article 13); and
- 'hosting' (Article 14).

These separate categories provide for increasing levels of engagement with the content – that is they assume different levels of passivity at each type of intermediary activity. There is virtually no knowledge of content when the intermediary is a 'mere conduit' (which merely provides for the transmission of information in an unaltered form at the direction of a third party); hosting, involving the storage of third parties' information, requires the intermediary still to play a technical, automatic and passive role.[82] Platforms prioritise content and target users – can they be said to be neutral, or does the fact that many of these features are driven by automated systems mean that they survive the test set down by the directive? The categories of intermediaries enjoying (at least in principle) immunity are not defined, and do not fit new services well – how, for example, do we categorise live streaming? The eCommerce Directive deliberately left untouched the liability regime for hyperlinks and search engines (information location tools). The result was that member states developed different approaches.

Hosts, when they become aware of problematic content, must act expeditiously to remove it or lose immunity. National implementation and court practice differ between member states considerably when assessing actual knowledge. The approaches ranged from the situation where some member states required a formal procedure resulting in an official notification by authorities in order to assume actual knowledge of a provider, to the other end of the scale where others leave it to the courts to determine actual knowledge. Nonetheless, according to the Court of Justice, awareness refers to facts or circumstances on the basis of which a 'diligent economic operator' should have identified the illegality in question.[83] It is most likely that social media platforms would fall within Article 14 e-Commerce Directive if they can claim any form of immunity at all. The take down procedures are not set out in the e-Commerce Directive, though some case law has developed in this area about what is adequate. Again, there is some difference between Member States; the Commission has contemplated some further EU level action in this regard. Certainly, more clarity about what is expeditious would be helpful.

Note there are special rules for intellectual property infringement too which affect intermediaries' duties. In addition to the 2004 Enforcement Directive[84],the European Parliament has recently agreed the proposal for a Directive on Copyright in the Digital Single Market.[85] It contains a controversial provision, Article 17 which excludes platforms that do not take diligent steps to ensure that pirate material is not displayed are excluded from the safe harbour provisions.

Impact on Regulatory Regimes

How would a regulatory regime fit with the immunity required by the e-Commerce Directive? Of course, this depends on the nature of the regime but a couple of general points can be made. Most notably, the logic of the directive is not to exclude ISS providers who provide hosting services from all forms of regulation. Indeed, they are not immune to all forms of legal action – for example, Article 14 immunity is not a response to an injunction. The provision relates to liability "for the information stored" and not other forms of possibility exposure to liability. This means that there is a difference between rules aimed at the content (which insofar as they are acceptable from a human rights perspective would in principle impose liability on the user unless the ISS host (a) was not neutral as to the content; and/or (b) did not take it

down expeditiously) and those aimed at the functioning of the platform itself (which might include rules as to how fast those systems should take content down). Indeed, the e-Commerce Directive recognises that some such rules could be imposed: recital 48 refers to the possibility of Member States imposing duties of care on hosts to 'detect and prevent certain types of illegal activities'. The placement of the recital suggests that it is aimed to clarify the meaning of the prohibition in Article 15 on Member States from requiring ISS providers to carry out general monitoring; recital 47 also clarifies that Article 15 does not concern monitoring obligations in a specific case. The boundary between these specific obligations and general monitoring is not clear. It would seem, however, that other techniques – specifically those that prevent harmful activity in the first place – would not be caught. It also seems that other pieces of legislation (e.g. the new copyright directive[86] already noted and the proposal on terrorist content on line[87]) are cutting down the scope of the Article 15 prohibition; where the limits of this gradually increase in monitoring lies especially in the context of the right to privacy is uncertain. The question, however, lies outside the scope of this report.

Audiovisual Media Services Directive (AVMSD)

The AVMSD is mainly targeted at television and analogous on-demand services. In the most recent revision of the directive, its scope has been extended and two new provisions dealing with 'video sharing platforms' (VSPs) have been introduced. While the directive does not say so, the effect of this amendment is to introduce these special rules on top of the e-Commerce Regime. Article 28b of the amended AVMSD provides that VSPs should be required to "take appropriate measures" to protect users from certain types of content. This is significant from two perspectives:

- It makes clear that Member States are obliged to regulate in this field; and
- Insofar as the directive expressly recognises the importance of freedom of expression, suggests where a boundary between conflicting interests might lie.

Article 28b AVMSD is a minimum harmonisation provision; Member States may impose stricter rules on VSP providers established on their respective territories. The assessment of whether provisions are appropriate will fall to the designated independent regulator in each Member State. The directive must be implemented by 19 September 2020.

Impact on Regulation

The revised AVMSD provides a partial approach to regulation of social media. It is partial because it is a minimum harmonisation measure but also because it applies only to VSPs as defined in the revised AVMSD. The meaning of VSP is complex and likely to be highly contentious[88] but one thing is clear – the provisions do not cover all social media platforms. A boundary will need to be drawn through that field, which may be understood as lying in different places across the various Member States. The outcome is, however, that as far as regulation of social media platforms lying within the definition of VSP member States will no longer have a choice as to whether to regulate. The question will be how they regulate and the extent to wish they each wish to go beyond the minimum set out in the revised AVMSD. To avoid the difficult boundary issues introduced by the choice to regulate just VSPs, it may make more sense for national rules to regulate social media in its entirety but notifying the Commission in accordance with the requirements of the e-Commerce Directive.

5. What Can We Learn from Other Models of Regulation?

Assuming that some sort of regulation (or self- or co-regulation) is necessary to reduce harm, what form should it take? The regulatory environment provides a number of models which could serve as a guide. In adopting any such models, we need, at least until Brexit, to be aware of the constraints of the e-Commerce Directive and the AVMSD. We also need to be aware of the limitations on governmental action arising from human rights considerations, specifically (though not limited to) freedom of expression. Limitations on rights must meet certain requirements, notably that they be proportionate. As we have shown, neither of these regimes foreclose regulatory activity entirely.

British and European countries have adopted successful regulatory approaches across large swathes of economic and social activity, for example broadcasting, telecommunications, data, health and safety, medicine and employment. We judge that a regulatory regime for reducing harm on social media can draw from tried and tested techniques in the regulation of these fields. This chapter provides a short overview of a range of regulatory models currently being deployed in these different sectors and identify regulatory tools and approaches such as risk assessments, enforcement notices etc within these that may have some relevance or applicability for a harm-based regulation model for social media. We discuss the regulatory frameworks frequently cited in relation to social media services, namely the electronic communications and broadcasting sectors which represent the transmission/intermediary and content contexts respectively. We also consider data protection, as data processing is at the heart of social media services. Given our view that social networks have strong similarities to public spaces in the physical world, we have also included some other regimes which relate to the safeguarding of public or semi-public spaces. Harm emanating from a company's activities has, from a micro-economic external costs perspective, similarity to pollution and we also discuss environmental protection

Telecommunications

Given that social media platforms are mainly not content providers but rather channels or platforms through which content is transferred from one user to another, one starting point is to look at the regulatory context for other intermediaries who also connect users to content: those providing the telecommunications infrastructure.

Electronic communications systems have long been subject to regulation.[89] The relevant rules are currently found in the Communications Act 2003[90]. The act implements the EU regime on telecommunications; the e-Commerce Directive (including its prohibition on licensing regimes) does not apply. There is, as a result of the EU requirements[91], no prior licence required under the Communications Act to provide electronic communications services (a change from the previous regime in the Telecommunications Act 1984), but a person providing a relevant "electronic communications network" or "electronic communications service" must under section 33 of the Communications Act give prior notification to OFCOM, the independent sector regulator.[92] Some services may still require licences – for example mobile operators in relation to spectrum use. While any person may be entitled to provide a network or services, that entitlement is subject to conditions with which the provider must comply – so the system here envisages compliance with standards.

The conditions comprise "general conditions" and "special conditions". As the name implies, "general conditions" apply to all providers, or all providers of a class set out in the condition. OFCOM may review the general conditions from time to time, in response to current priorities. The most recent version came into

force 1 October 2018. Special conditions apply only to the provider(s) listed in that special condition (see section 46 Communications Act). The conditions are set by the regulator in accordance with the act.

Section 51 sets down matters to which general conditions may relate. They include:

> *"conditions making such provision as OFCOM consider appropriate for protecting the interests of the end-users of public electronic communications services"*

which are elaborated to cover matters including the blocking of phone numbers in the case of fraud or misuse, as well as, in section 52, the requirement to have a complaints mechanism. This latter point is found in General Condition 14 (GC14)[93] which obliges communications providers to have and to comply with procedures that conform to the OFCOM Approved Code of Practice for Complaints Handling[94] when handling complaints made by domestic and small business customers. This may be a point that could be copied over to a social media regime. General conditions also cover public safety (in relation to electro-magnetic devices). The specific conditions mainly relate to competition in the telecommunications market, specifically the rights of businesses to have access to networks on fair, reasonable and non-discriminatory terms.

In terms of enforcement, section 96 gives OFCOM the power to impose fines on providers for non-compliance with the conditions. Ultimately, OFCOM has the power to suspend the entitlement to provide the service (see section 100).

Broadcasting

OFCOM also has regulatory responsibilities in relation to broadcasters and those providing on demand audiovisual services, also found in the Communications Act. The Communications Act is the most recent in a long line of statutes which impose detailed obligations on broadcasters.

The current regime distinguishes between those transmitting on terrestrial capacity, and who are public service broadcasters, and those providing content running across the various means of distribution. While the terrestrial licences are limited, licences in relation to content services are not – they are awarded on receipt of an application with the appropriate fee (currently £2,500 per application). Award of a licence is not automatic – licence holders must be "fit and proper" persons and certain groups (e.g. political parties) are not permitted to hold a licence. This is a distinctive aspect of the broadcasting regime. Ownership restrictions on the accumulation of media interests also apply in addition to general competition policy; these sector specific rules have been whittled away over the years. Note that there is a separate public interest test which may apply to media mergers, to ensure that the public interest in having a diverse range of voices represented in the media is taken account of.

Licensees must comply with content standards as set out in the Communications Act; the licensee must also demonstrate it has compliance procedures in place – for example, there must be enough staff of sufficient seniority trained to understand the licence conditions. This is a requirement that OFCOM regularly enforces. In this context, it is also worth noting that OFCOM has a responsibility to develop a code with regard to acceptable content (in accordance with section 319 Communications Act) and as part of that process gives guidance to broadcasters as to the boundaries between acceptable content, as well as how to respect other obligations (e.g. impartiality on the news and current affairs programmes). This suggests that it is possible for an independent regulator to develop more specific rules in difficult to define areas, based on broad principles in statute.

Digital Economy Act 2017

The Digital Economy Act 2017[95] covers a range of topics – we will focus on just one aspect: the provisions in relation to age verification and pornography which is found in Part 3 of the Act[96]. While we note the Government's expectation that social media companies would take action to protect their users in line with the Social Media Code of Practice[97], we do not think that a voluntary code is sufficient; the content of the code (which is principles based) may prove a starting point for elaborating the harms which social media companies should tackle).

The obligation is to ensure that pornographic material is not made available online to people under 18 – essentially this requires age verification. The operator has freedom in how to attain this goal, but the Age Verification Regulator may check these steps. It may also issue guidance as to how age verification may be carried out[98]. It may issue enforcement notices and/or impose penalties if a person has failed in this duty (or refuses to give the regulator information requested). The Digital Economy Act also empowers the regulator to issue notices to others who are dealing with the non-complying operator (section 21), such as credit card or other payment services.[99] According to the Explanatory Memorandum[100], the purpose of these provisions is "to enable them to consider whether to withdraw services". This is an illustration of mechanisms that may be used in relation to companies that are not directly within the scope of regulation, notably because they are established overseas. In relation to extreme pornography only, the regulator has the power to request that sites are blocked.

By contrast to the regimes for telecommunications, broadcasting and data protection, the Digital Economy Act does not specify that the age verification body must be independent.

Data Protection

Another possible model in the sphere of information technology is that of data protection and specifically, the General Data Protection Regulation (GDPR), which was implemented in the UK in May 2018 through the Data Protection Act 2018. The provisions dealing with the independent regulatory authority, an essential part of the regime, and the minimum standards of independence are set down in the GDPR; for the UK, these are set out in relation to the Information Commissioner and her office (ICO) in the Data Protection Act 2018[101].

The current regime requires those processing personal data to register with the ICO to renew the registration annually, and to pay a fee (which varies depending on the size and nature of the organisation). Failure to do so is a criminal offence. The GDPR removes the annual renewal obligation but data protection fees still apply, by virtue of the Digital Economy Act 2017. Some information (e.g. name, contact details) will have to be submitted with this fee, but the current notification regime which required details about the data processing will cease to exist.

Central to the GDPR is the principle of accountability which can require a data controller to show how it complied with the rules. These essentially put obligations on controllers to process data in accordance with the 'data processing principles' set down in Article 5 GDPR.

Another theme is a form of precautionary principle in that data controllers must comply with both privacy and security by design. This could be described as a risk-based approach, as can be seen in the requirements regarding data security. For example, data controllers are required to "ensure a level of data

security appropriate to the risk" and in general they should implement risk-based measures for ensuring compliance with the GDPR's general obligations. Controllers should ensure both privacy and security by design. High risk processing activities trigger the need for a privacy impact assessment (PIA) to be carried out (Article 33 GDPR). Article 34 specifies that where the PIA suggests that there is a high risk, the controller must ask the supervisory authority before proceeding.

In addition to the requirements of the GDPR, the Data Protection Act 2018 introduced a requirement for the ICO to consult on an age appropriate design code in relation to information society services that are likely to be accessed by children.[102] In drafting the code, the ICO must have regard to the best interests of the child, as well as regard to the UN Convention on the Rights of the Child. At the time of writing, the ICO had not yet issued the code. This provision highlights the fact that design is not neutral, and that some design choices may be better or worse for particular groups of users. Indeed, in the Lords debate discussing the amendment, Baroness Kidron emphasised that this would be the first time that the ICO would have to consider design choices aimed at extending user engagement – "that is, those design features variously called sticky, reward loops, captology and enrapture technologies that have the sole purpose of making a user stay online. These will be looked at from a child development point of view".[103]

As regards enforcement, the themes of risk-based regulation and 'by design' privacy and security choices feed into the assessment of the size of fines to impose on a controller (or processor) in breach under the GDPR, as the authority will have "regard to technical and organisational measures implemented" by the processor. The ICO has set out its strategy for enforcement in its Regulatory Action Policy, in which it identified its 'hierarchy of regulatory action' which reflects this theme.[104] The regime also allows for the serving of enforcement notices.[105] The regime allows the ICO to encourage compliance, through information and advice, as well as enforce compliance through sanctions.

The Data Protection Act is an important model also in terms of enforcement powers: the ICO's powers were strengthened significantly under the DPA18 by comparison with the earlier act. For example, while the 1998 Act gave the ICO the power to serve an information notice,[106] the DPA18 makes it an offence to make a false statement in response to such a notice. The ICO's may also serve an 'urgent' information notice, requiring a response within 24 hours. The ICO may serve an assessment notice,[107] giving the ICO access to premises and specified documentation and to interview staff – the urgent assessment notice gives very little notice to its recipient

Note that individuals also have a right to bring actions for data protection failings; Article 80(1) GDPR also envisages the possibility of mandated civil society organisations acting on behalf of the individuals affected by non-complaint processing. Article 80(2) GDPR also states:

Member States may provide that any body, organisation or association referred to in paragraph 1 of this Article, independently of a data subject's mandate, has the right to lodge, in that Member State, a complaint with the supervisory authority ….

The Data Protection Act 2018 provides that the Secretary of State must review the UK's provisions for the representation of data subjects under Article 80, including whether to exercise the power under Article 80(2), but for the time being no bodies have been given an independent right to complain. This decision has been criticised as constituting a weakness in collective redress systems, which are particularly important in contexts where many people may suffer harms the loss in respect of which is hard to quantify.[108]

Health and Safety

Another model comes from outside the technology sector: health and safety. This is particularly relevant if we move, as suggested, beyond the binary consideration of whether social media platforms are either publishers or part of the transmission network but rather view them as constituting public or quasi-public environments.

Arguably the most widely applied statutory duty of care in the UK is the Health and Safety at Work Act 1974 (HSWA 1974)[109] which applies to almost all employers and the myriad activities that go on in them. The HSWA 1974 does not set down specific detailed rules with regards to what must be done in each workplace but rather sets out some general duties that employers have both as regards their employees and the general public.

It elaborates on particular routes by which that duty of care might be achieved: e.g. provision of machinery that is safe; the training of relevant individuals; and the maintenance of a safe working environment. The Act also imposes reciprocal duties on the employees.

While the HSWA 1974 sets goals, it leaves employers free to determine what measures to take based on risk assessment. Exceptionally, where risks are very great, regulations set down what to do about them (e.g. Control of Major Accident Hazards Regulations 1999[110]). In respect of hazardous industries, it may operate a permission regime, in which activities involving significant hazard, risk or public concern require consent; or a licensing regime to permit activities, such as the storing of explosive materials, that would otherwise be illegal.

The area is subject to the oversight of the Health and Safety Executive (HSE), whose functions are set down in the Act. It may carry out investigations into incidents; it has the power to approve codes of conduct. It also has enforcement responsibilities and may serve "improvement notices" as well as "prohibition notices". As a last measure, the HSE may prosecute. There are sentencing guidelines which identify factors that influence the heaviness of the penalty. Points that tend towards high penalties include flagrant disregard of the law, failing to adopt measures that are recognised standards, failing to respond to concerns, or to change/review systems following a prior incident as well as serious or systematic failure within the organisation to address risk.

The HSWA regime was subject to review due to concerns about a 'compensation culture' as well as the complexity of the regulations that support the general system and the codes of practice approved by the HSE. Although the review found some of the concerns regarding complexity may have been justified, overall the review found that there was no case for radically changing the system. Specifically, the "'so far as is reasonably practicable' qualification in much of health and safety legislation was overwhelmingly supported by those who responded to the call for evidence on the grounds that it allows risks to be managed in a proportionate manner"[111]. Insofar as there were weaknesses in the system, it arose from businesses misunderstanding the regulatory requirements.

Environmental Protection

Another regime which deals with spaces is the Environmental Protection Act 1990[112]. It imposes a duty of care on anyone who produces, imports, keeps, stores, transports, treats or disposes of waste and brokers or those who control waste ("waste holders") (section 34), as well as on householders (section 75(5)), but

does not state the person to whom the duty is owed. Waste holders must register with the possibility of a fine for non-compliance; there is a prohibition on unauthorised disposal of waste backed up with a criminal penalty.

More detail on what the duty of care requires is set down in secondary legislation and codes of practice give practical guidance. As regards waste holders, they are under a duty to take all reasonable steps to: prevent unauthorised or harmful deposit, treatment or disposal of waste;

- prevent a breach (failure) by any other person to meet the requirement to have an environmental permit, or a breach of a permit condition;
- prevent the escape of waste from their control;
- ensure that any person you transfer the waste to has the correct authorisation; and
- provide an accurate description of the waste when it is transferred to another person.

The documentation demonstrating compliance with these requirements must be kept for two years. Breach of the duty of care is a crime.

Householders' duties are more limited: they have a duty to take all reasonable measures to ensure that any household waste produced on their property is only transferred to an authorised person – a householder could be prosecuted for fly-tipping of waste by a contractor (plumber, builder) employed by the householder.

As well as this duty of care, businesses are required under Reg 12 of the Waste (England and Wales) Regulations 2011[113] to take all such measures as are reasonable in the circumstances to:

- prevent waste; and
- apply the "waste hierarchy" (which is a five step strategy for dealing with waste ranging from prevention through recycling to disposal which derives from the EU Waste Framework Directive (2008/98/EC)[114]) when they transfer waste.

In doing so, business must have regard to any guidance developed on the subject by the appropriate authorities.

The responsible regulators are the Environment Agency/Natural Resources Wales/Scottish Environment Protection Agency and local authorities. They may issue enforcement notices, and fines may be levied. If criminal action is taken, there is a sliding scale based on culpability and harm factors identified in guidance. The culpability assessment deals with the question of whether the organisation has deliberately breached the duty, done so recklessly or negligently – or to the contrary, not been particularly at fault in this regard.

Assessment of comparative regimes

The sectors discussed above operate under general rules set by Parliament and refined by independent, evidence-based regulators and the courts in a transparent, open and democratic process. Modern effective regulation of these sectors supports trillions of pounds of economic activity by enforcing rights of individuals and companies. It also contributes to socially just outcomes as intended by Parliament through the internalisation of external costs and benefits.

There are many similarities between the regimes. An important point is that most of the regulators are specified to be independent. That is, once the general objectives have been set down by Parliament, the regulators are not under instruction from parliament as to how they carry out their statutory obligations. This insulates the operation of the regulatory systems from political interference. While this is important in all sectors, it is especially so where human rights, specifically freedom of expression, are in issue. It is also important that the regulator is independent from industry influence too. Members of such bodies must all comply with Nolan Principles, which also support independence. One question that is not addressed in any of the regulators discussed is the question of diversity of the governing body. Certainly – and by contrast to systems on the continent – British regulators have tended to take an approach according to which those who regulate are seen as experts and not representatives of the population which may explain why hitherto that have been no rules about the diversity of governing bodies.

Another key element of many of the regulators' approach is that changes in policy take place in a transparent manner and after consultation with a range of stakeholders. Further, all have some form of oversight and enforcement – including criminal penalties. Matters of standards and of redress are not left purely to the industry, though the rights of individuals under a number of these systems (e.g. telecommunications and health and safety at work) are limited.

There are, however, differences between the regimes. One point to note with regards to the telecommunications and the broadcasting regimes is that in both instances OFCOM may stop the provider from providing the service. While the data protection regime may impose – post GDPR – hefty penalties, it may not stop a controller from being a controller. Again, with regard to HSE, particular activities may be the subject of a prohibition notice, but this does not disqualify the recipient from being an employer. The notice relates to a particular behaviour. We question whether this is appropriate in the light of freedom of speech concerns in our discussion of regulatory sanctions in Chapter 9. Similar concerns may relate to 'fit and proper' tests – their use in the context of broadcasting was justified by the limited access to spectrum and the privileged position that broadcasters had in selecting information to be conveyed and thereby influencing public opinion.

Another key difference between the telecommunications (especially broadcasting) regimes and the others is they specify the standards to be met in some detail. A broadcasting approach in particular – which looks to individual items of content – would be difficult to scale, as has been well recognised. We also question whether the degree of specificity means that the regimes are too tightly orientated to a particular type of service. For the other regimes, although there are general obligations identified, the responsibility in both instances lies on the controller/employer to understand the risks involved and to take appropriate action, though high-risk activities in both regimes are subject to tighter control and even a permissioning regime. Allowing operators to make their own assessment of how to tackle risks means that solutions may more easily keep up with change, as well as be appropriate. This allows for a certain amount of future proofing.

A risk-based approach could also allow the platforms to differentiate between different types of audience – and perhaps to compete on that basis

There are perhaps techniques from individual regimes that are worth highlighting and which should be borne in mind when seeking to develop a regime:

- the data protection and HSE regime highlight that there may be differing risks and that has two consequences:

  ° that measures should be proportionate to those risks; and

  ° that in areas of greater risk there may be greater oversight;

- the telecoms regime emphasises the importance of transparent complaints mechanisms – this is against the operator (and not just other users);

- the broadcasting regime envisages that a regulator may develop more detail on content standards, useful in identifying the boundary between challenging content and that which is impermissible – the HSWA also envisages the development of codes to guide business and as the Löfstedt Review emphasised the codes should be written in clear and non-technical language;

- the broadcasting regime also illustrates the role of focussing on processes – notably the number of staff that are trained to deal with compliance issues;

- the environmental regime introduces the ideas of prevention and prior mitigation, as well as the possibility for those under a duty to be liable for the activities of others (e.g. in the case of fly-tipping by a contractor); and

- the Digital Economy Act has mechanisms in relation to effective sanctions when the operator may lie outside the UK's jurisdiction.

Outline of a proposed regulatory model

We have suggested that an appropriate analogy for social media platforms is that of a public space. The law has proven very good at this type of protection in the physical realm. Workspaces, public spaces, even houses, in the UK owned or supplied by companies have to be safe for the people who use them. One technique used is that where the law imposes a statutory duty of care on the owners of those spaces as we have discussed the HSWA. We take this approach as our starting point for a new regime.

Statutory duties of care can be expressed in terms of what they want to achieve – a desired outcome (i.e. the prevention of harm) rather than necessarily regulating the steps – the process – of how to get there. This fact means that duties of care work in circumstances where so many different things happen that you would be unable to write rules for each one. This generality works well in multifunctional places like houses, parks, grounds, pubs, clubs, cafes, offices and has the added benefit of being (as noted) to a large extent futureproof. By taking a similar approach to corporate owned public spaces, workplaces, products etc in the physical world, harm can be reduced in social networks. Making owners and operators of the largest

social media and messaging services responsible for the costs and actions of harm reduction will also make markets work better.

In the new regime, as with the HSWA, we envisage that the initial responsibility lies with the company, which must take reasonable measures to prevent foreseeable harm – and we envisage the harms being specified (at a high level) by statute. Essentially, this is an approach tied to outcomes-based accountability.[115] We envisage the need for a regulator, as used in most of the regimes we have discussed and that the regulator must be independent. So, while the company has freedom to adopt its own approach, the issue of what is 'reasonable' is subject to the oversight of a regulator. Part of the oversight regime involves the measuring of progress towards the driving down the risk and/or incidence of harm through the harm reduction cycle. Further questions arise – such as how users' rights are to be taken account of, what sorts of sanctions should be made available, and how will companies identify harms – and for this, we return to the points highlighted from the other models as we set out below.

6 The Statutory Duty of Care

The idea of a duty of care derives from the tort of negligence.[116] It provides an individual with a right to recompense; torts in general are not designed to be punitive or impose regulatory standards.[117] There are four elements which need to be satisfied for a claimant to successfully claim negligence:

- the defendant owed the claimant a *duty of care* to avoid the injury of which the claimant claims[118];
- the defendant *breached* the duty of care by behaviour that does not reach the standard of reasonable care;
- the defendant's breach *caused* the damage; and
- the claimant's damage was not *too remote/unforeseeable*.

Statutory duties of care were established in contexts where the common law doctrine seemed insufficient. In some instances this was because the relationship between the victim and defendant was not sufficiently close to impose a duty of care (e.g. Occupiers Liability Act 1957[119]). Further, the duty of care only arises in relation to certain types of harm – the courts have been reluctant to admit mental injury short of a recognised psychiatric illness,[120] which can lead to significant hardship being outside the scope of the remedy. The case law is particularly complex where a victim seeks to claim that the defendant is liable for the actions of third parties. Although the courts have not excluded this possibility[121], the cases in which this argument is successfully pleaded are rare. We therefore propose – as in the HSWA 1974 – a *statutory* duty of care which can clarify the existence of the duty, the types of harms to be protected against and provide a system to clarify the level at which the duty has been breached. In such a general system, the issue of causation in the instance of a particular individual falls away. In sum, a statute can be used to cure or ameliorate the deficiencies of the common law.

A statutory duty of care is straightforward in principle. A person (including companies) specified to be under a duty of care must take care in relation to a particular activity, usually as it affects particular people or things. If that person does not take care and others come to harm as a result then there are legal consequences. A duty of care does not require absolute protection from harm – the question is

whether sufficient care has been taken. There is thus a distinction between term 'reasonably practicable' and (physically) possible/practicable. 'Reasonably practicable' is a narrower term. If it can be shown that there is a gross disproportion between the cost of a measure and its benefit, the risk being insignificant in relation to the sacrifice, the person upon whom the obligation is imposed would discharge the onus which is upon him. The foreseeability of a risk is a relevant factor too. So – looking at existing case law - an employer would not in breach of a duty of care in failing to take precautions against an unknown danger.[122]

For statutory duties of care, while the basic mechanism may in each instance be the same, the details in each statutory scheme may differ – for example, the level of care to be exhibited and the types of harm to be avoided. Even within the same context, duties could be differently phrased. We note that the Children's Commissioner has developed a draft duty of care and that the NSPCC, as well as the significant work it has done on identifying harms to children online, has produced a regulatory model working with Herbert Smith Freehills (based upon our duty of care work)[123]. A key difference is that our work is broader in scope; the children's charities understandably focus on the well-being of children.

Our starting point is the HSAW74, which as we have noted applies to almost all workplaces in the UK. It sets out the primary statutory duty of care in S2(1):

> *'It shall be the duty of every employer to ensure, so far as is reasonably practicable, the health, safety and welfare at work of all his employees.'*

The Occupiers Liability Act 1957 which applies to almost all private land in the UK but which does not form part of a regulatory scheme phrases its statutory duty of care in S2(2)

> *'The common duty of care is a duty to take such care as in all the circumstances of the case is reasonable to see that the visitor will be reasonably safe in using the premises for the purposes for which he is invited or permitted by the occupier to be there.'*

Drafting a precise duty of care for social media and messaging services would require more detailed work but a starting point could be:

> *'It shall be the duty of every qualifying operator:*
>
> *to ensure, so far as is reasonably practicable, that the users of their service are free from harm arising from its operation or use;*

It seems that while we envisage a significant role for a regulator in determining where there is a risk of harm and reasonable solutions to address that harm, it could be that the statute identifies – as do the HSWA 1974 and the Environmental Protection Act 1990 – particular steps an operator should address.

The following sections detail how we think these points should be addressed, as well as outlining how the regulatory scheme would work and the nature of the regulator.

Harm to people who are not users

Our original proposal limited harms to the users of the qualifying services to harms on those services. However, we note that:

a) harm may implicate more than one platform and, more generally,

b) people are harmed by content on social media and messaging services when they themselves are not customers of those services.

As regards (a): we note that Twitch, for instance, is already grappling with this third-party service problem. After user feedback, Twitch gave itself powers[124] to sanction customers who use another service (Twitter, say) to organise attacks on a fellow Twitch user. Twitch extends this to IRL meet-ups. Twitch requires evidence to be presented to it. This suggests that a provider's responsibility does not end with the limits of its own platform and that deliberate offenders will move from platform to platform. We note that the process of regulation could bring service providers of all types together to share knowledge about harms within and between platforms, putting commercial interests to one side[125].

As regards (b): consider the harm suffered by a woman who has revenge porn posted on a service of which she is not a customer. The service provider's obligation to the victim should not depend on whether or not she had signed up to the service that was used to harass her. Extending the statutory duty to individuals who are not users of the service is important as it is far from certain that, under the common law duty of care, a duty would arise to such an individual; and, given the lax enforcement of the criminal law, it is unlikely that the existence of the criminal offence has much deterrent effect. Any extension of the scope of the duty would continue to be subject to a reasonableness test.

We feel that the structure as used in the Environmental Protection Act may be too broad as it seemingly does not limit the extent of the duty to a particular group. While this is understandable in the environmental context, we are cautious about imposing it here. Another approach might be that already used by the HSWA 74 to tackle harm to people outside the immediate duty of care. Section 2 of the Act[126] covers the relationship between employer and employee (ie a contractual relationship akin to the relationship between platform and user which is governed by terms of service). But Section 3 is wider. It provides that:

It shall be the duty of every employer to conduct his undertaking in such a way as to ensure, so far as is reasonably practicable, that persons not in his employment who may be affected thereby are not thereby exposed to risks to their health or safety.

The connecting factor in the HSWA74 is whether a person 'may be affected': a very broad category which, were it to be applied analogously to the online context, could fill the gap in protection as we have attempted above. We note that the HSWA74 provides a lesser protection to third parties than it does to employees and have adapted this principle accordingly. We therefore propose that the statutory duty have a second element:

b) to conduct its undertaking so that, so far as is reasonably practicable, people who may be affected by the service and are not users of that service are not appreciably harmed as a result of its operation or use'.

7 Which social media services should be regulated for harm reduction?

We set out above a proposed system where every company that operates a qualifying social media or messaging service used in the UK is subject to some general rules or conditions, notably a duty of care to their users. In this chapter, we discuss which social media services would be subject to such a statutory duty of care towards their users. Some definition is required as although terms such as intermediary, platform or social media are often used, there is no common understanding of these terms even when used in statute.[127] For example, as Bunting noted, there is no legislative definition of 'social media platforms' although this is the term used in the Digital Economy Act.[128]

Qualifying social media services

We suggest that the regime apply to social media services used in the UK that have the following characteristics:

- Have a strong two-way or multiway communications component;
- Display and organise user generated content publicly or to a large member/user audience;
- Are not subject to a detailed existing content regulatory regime, such as the traditional media (broadcast or press).

Our proposals do not displace existing laws – for example, the regulatory regime in relation to advertising overseen by the Advertising Standards Authority (ASA[129]) whose rules are pertinent here but which relate to the content of the advertisement, nor the competition and consumer protection regimes, recently the subject of proposed reform in respect of digital and other issues.[130]

A regulator would produce detailed criteria for qualifying social media services based on the above and consult on them publicly. The regulator would be required to maintain a market intelligence function to inform consideration of these criteria. Evidence to inform judgements could come from: individual users, civil society bodies acting on behalf of individuals, whistle-blowers, researchers, journalists, consumer groups, the companies themselves, overseas markets in which the services operate, as well as observation of trends on the platforms.

In order to maintain an up to date list, companies established within the jurisdiction which fall within the definition of a qualifying social media or messaging service provider would be required in law to notify the regulator after they have been operating for a given period. Failure to do so would be an offence. Notification would be a mitigating factor should the regulator need to administer sanctions.

Providing a future-proof definition of a qualifying social media service is tricky. However, we feel that giving the regulator freedom from political interference to make a list allows for some developments in approach in the light of technological and market changes rather than defining it in detail in legislation. The regulator making this list also reduces the risk of political interference – it is quite proper for the government to act to reduce harm, but in our view there would be free speech concerns if the government was to say who was on the list – potentially favouring some views over others. If there was a concern about the legitimacy of this process (which is unlikely given the responsibilities of OFCOM in relation to broadcasting), an alternative would be for the regulator to advise the Secretary of State and for them

to seek a negative resolution on the list in Parliament but, in our view, this starts to introduce a risk to independence and freedom of speech.

Services Within Scope

Whilst our proposed definition catches large services such as Facebook, Twitter, YouTube, LinkedIn, TikTok, Kik etc there are a range of other services also in scope that we discuss here.

Internet forums have some of the characteristics we set out above. In a risk-based regime (see below) many would be deemed low risk and unlikely to be much affected. We do not intend to capture blog publishing services, but it is difficult to exclude them from scope when applying general characteristics of the sort we have identified.

Gaming and messaging have become closely entwined, both within games and using services outside games for in game discussion, viewing and commentary. Game video streaming service Twitch has had well-documented abuse problems[131] and has arguably more sophisticated banning regimes for bad behaviour than other social networks. Twitch allows gamers to stream content that the gamers have generated (on games sites) with the intention of interacting with an audience about that content. Twitch provides a place for that display, multiway discussion about it and provides a form of organisation that allows a user to find the particular content they wish to engage with. Discord a messaging service for gamers now with over 140 million users 'almost quadrupled' in size during 2017-18. Following reports of harm Discord is reported to be increasing its focus on user safety[132]. Discord is an example of a third party messaging service that runs outside of a game to provide real-time in game messaging between participants and others. Services of this nature would be covered by our proposal. Other gaming services with a strong social media element should also, particularly those with large internal messaging services and a strong youth user base.

Services do not need to include (much) text or voice: photo sharing services such as Pinterest could fall within the regime too.

'Messaging' services

Messaging services are not necessarily private and also give rise to risks to individuals. We encountered disturbing reports of harms arising in messaging services, for instance:

> *'One teen chat app has featured in more than 1,100 child sexual abuse cases in the last five years, the BBC has found. Of 29 police forces that supplied information to the BBC, all but one had child exploitation cases involving Kik'*[133].

And

> *'Images of child sexual abuse and stolen credit card numbers are being openly traded on encrypted apps, a BBC investigation has found. Security experts told Radio 4's File on 4 programme that the encrypted apps were taking over from the dark web as a venue for crime. The secure messaging apps, including Telegram and Discord, have become popular following successful police operations against criminal markets operating on what is known as the dark web - a network that can only be accessed by special browsers.'* [134]

Although some messaging service providers do carry out pro-active moderation[135], at least of unencrypted parts of their services, it is questionable if this is enough. Insofar as reasonably foreseeable harms arise, they should be risk managed by service providers. We continue to take the view that private communication, for which the model in Article 8 ECHR is essentially one-to-one communication, lies outside our proposed regime.

In the last year, it has become clearer that messaging services have gone beyond small groups supporting existing relationships – familial, friendship or work. We now observe a trend towards large groups and groups becoming findable to non-members who can join if there is room in the group[136]. The size of these groups suggests that the communication mediated via the service is neither private nor confidential. Other characteristics also indicate the non-private nature of the communication, notably the growing practice of public groups, sharing of group links and browsers and search apps for groups. Services that enable the creation of public groups and/or large groups would, in our view, become qualifying services under our proposal and fall under the statutory duty of care regime.

Reasonably foreseeable harms in a messaging service might be quite different to those in a public-facing social media service and may therefore require different responses. For instance, where the bulk of a service is not visible to the operator due to a business decision about encryption there should be a far more responsive and effective notice and remedy process for people in a group who have experienced harm. As we explain below, we propose a risk-managed harm reduction process which would lead to different measures to those for traditional social media.

Search engines

The government's broad definition of online harms led to us being asked whether a duty of care regime could apply to general search engines, of the likes of Google. In search, as in social media, the information presented to the user is a result of corporate decisions about terms of service, software and resources deployed to enforce and keep these up to date.

YouTube, the world's second biggest search engine, would be covered by our proposals and we have noted that its recommender algorithm (see Tufecki's critique[137]) is of particular concern.[138] Given that, can we continue to distinguish between social network sites and general search engines? There are indications that harm can arise through search engines: for example, Google is working on anti-radicalisation and other aspects of harm reduction in search.[139] We also note the disturbing research by Anti-Toxin for Tech Crunch into child abuse imagery on Bing[140], which Google had prevented returning in apparently identical searches presumably by better risk management. Consumers are not given information that labels one search engine as riskier than the other. In search, as in social media, the information presented to the user is a result of corporate decisions about service design, software and the resources the company chooses to deploy in maintaining and enforcing these.

On that basis, search engines should come into a risk-managed harm reduction framework. But is it this statutory duty of care framework? Search engines do not show the level of interaction between users that we had originally envisaged as a criterion for a qualifying service. Further, discoverability of information may raise a whole set of issues about public service and impartiality that may not be best considered through the lens of a statutory duty of care[141] and may raise questions about freedom of expression.[142] Finally, could a regulator manage search as well as social media? We can see arguments for including and excluding search from a duty of care regime but have not had time to consider the issues fully.

Should big and small services be regulated?

In our early thinking for this project we suggested that regulation should only apply to the largest social media services with over 1 million UK users. In the light of the government's approach, that

> *'legislation that will cover the full range of online harms, including both harmful and illegal content.'*[143]

we were asked by a range of stakeholders why we had proposed a more limited approach. Some suggested that a duty of care might work better in broader application; others that, for children in particular, the size of social network was immaterial – terrible harm could occur in only small networks. Responses to the ICO consultation on the Age Appropriate Design Code[144] also took this comprehensive approach to harm reduction for children.

We therefore came to a view that there should be no de minimis user/customer threshold for the duty of care. Some groups are sufficiently vulnerable (e.g. children) that any business aiming a service at them should take an appropriate level of care, no matter what its size or newness to market. Beyond child protection, basic design and resourcing errors in a growth stage have caused substantial problems for larger services.[145] Much of the debate on AI ethics attempts to bake in ethical behaviour at the outset.[146] The GDPR emphasis on privacy by design also sets basic design conditions for all services, regardless of size. We are struck that in other areas even the smallest businesses have to take steps to ensure basic safety levels – the smallest sandwich shops have to follow food hygiene rules. In both these cases, risks are assessed in advance by the companies concerned within a framework with a regulator.[147]

We note that Parliament made two major statutory duties of care we that we discuss above (Occupiers Liability 1959 and HSAW 1974) above apply almost pervasively, not substantially constrained by size of unit nor by a pre-assessment of the level of danger.

8 Definition of harms

When Parliament set out a duty of care it often sets down in the law a series of prominent harms or areas that cause harm that they feel need a particular focus, as a subset of the broad duty of care. They may link the harms to specific groups of persons to whom a duty of care is owed (as discussed above). This approach has the benefit of guiding companies and the regulator on where to focus and makes sure that Parliament's priorities are not lost. We envisage that more detailed guidance would be given in a code developed by the (independent) regulator and revised from time to time. We further anticipate that revisions should be evidence-based, include consultation with civil society, as well as surveys of user attitudes – in this aspect, process would reflect that adopted by a number of regulatory bodies within their respective fields of responsibility: OFCOM, ASA, BBFC. This approach provides some insulation from politically motivated interference with speech, allows a range of interests to be considered and provides flexibility for the code to be developed in response to changing circumstances.[148]

We propose setting out the key harms that qualifying companies have to consider under the duty of care, based in part on the UK Government's Internet Safety Green Paper and the principles underpinning the draft Social Media Code[149].

Harmful threats

This category includes any statement of an intention to cause pain, injury, damage or other hostile action such as intimidation. Psychological harassment, threats of a sexual nature, threats to kill, racial or religious threats known as hate crime. Hostility or prejudice based on a person's race, religion, sexual orientation, disability or transgender identity. Many of these acts would be likely to fall within the definition of one crime or other as currently drafted, though seemingly under-prosecuted. We would extend hate crime to include misogyny currently being reviewed by the government[150] and to disabled people[151].

Economic harms

We group under economic harms a range of activities where people seek to use online networks as a vector or medium to rip others off. The range of activity is very broad and at an exceptionally high level. The National Trading Standards eCrime team maintains a rolling list[152] of the latest online scams. Work by Anderson et al as long ago as 2012 estimated that:'cybercrime is now the typical volume property crime in the UK'[153]

The 2017 Public Accounts Committee report[154] into online fraud concluded that:

> *'Online fraud is now the most prevalent crime in England and Wales, impacting victims not only financially but also causing untold distress to those affected. The cost of the crime is estimated at £10 billion, with around 2 million cyber-related fraud incidents last year, however the true extent of the problem remains unknown. Only around 20% of fraud is actually reported to police, with the emotional impact of the crime leaving many victims reluctant to come forward.'*

Which? has also uncovered evidence that more than 96% of frauds reported to Action Fraud (a branch of the City of London police) are not solved. In supplementary evidence to the Lords Communications Committee inquiry on Internet Regulation, they put forward a number of reasons for the poor resolution of fraud cases, including the "invisible" nature of the offenders, given that the ONS estimates that 55% of frauds have a digital element.

> *"Even with cutting-edge technologies to track down online or telephone fraudsters, investigators have an uphill battle compared with detectives working on a physical crime where there is CCTV evidence, eyewitnesses sightings etc".*[155]
>
> *Other elements of consumer detriment that Which? have identified online include: unsafe products being sold by online retailers; fake reviews and the impact of data collection and use on competition and choice for consumers, including through personalised pricing, targeting and digital advertising*[156].

Economic harm encompasses intellectual property crime. The UK government's IP crime and enforcement report 2017/18 reports social media as the second highest location for IP crime tackled by Trading Standards:

Trading Standards statistics*

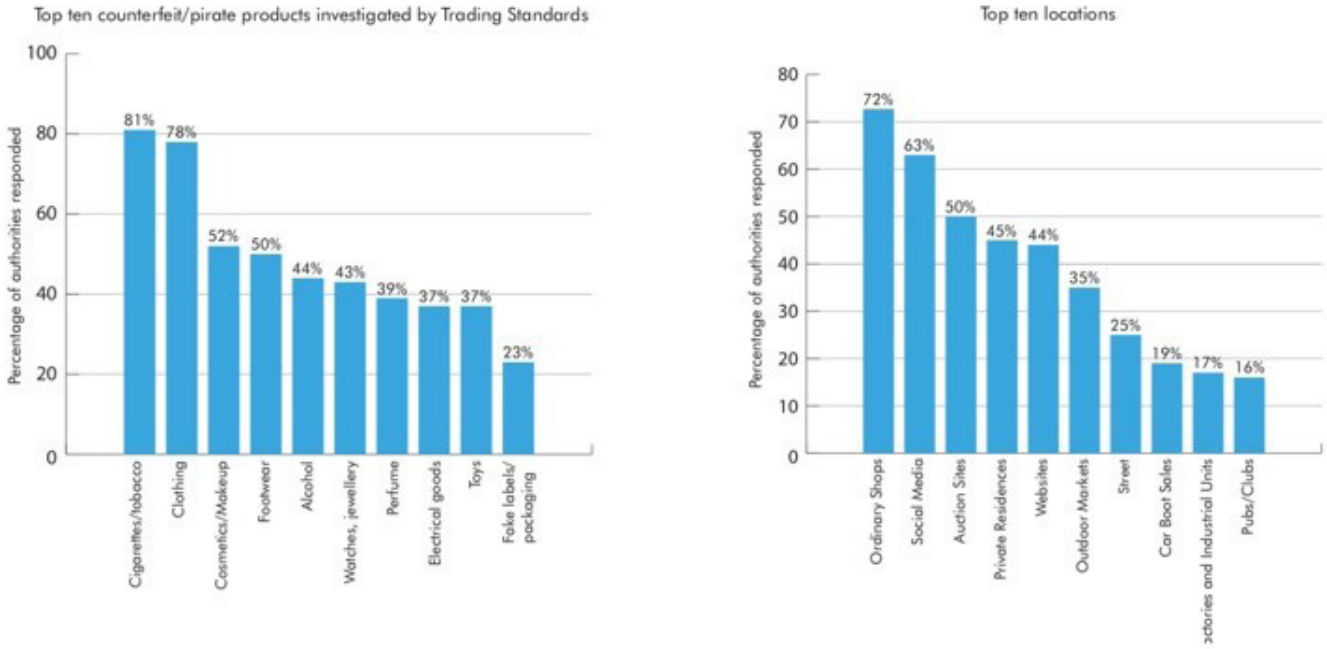


Figure 1 IP crime types and locations – HMG Trading standards successes IP crime and enforcement report 2017 to 2018

The report[157] continues:

> *'Work by the NMG and partners shows that illicit traders set up bogus Facebook accounts using closed groups and operate within 'local selling group' to attract customers who are often unaware that the products they are buying are counterfeit. Counterfeiters using these online platforms are also engaged in forgery of passports, driving licenses and other official documents as well as the use and supply of controlled drugs, weapons and other illicit trade.'*

The music, sports etc ticket reselling industry has attracted much criticism for sharp practices, with the ASA and CMA taking part in concerted action. However, people use online platforms to find tickets in the first place. In their March 2019 report into live music[158], the Commons DCMS Select Committee said:

> *'…media owners also have a responsibility to the audiences they serve. Google has repeatedly allowed ticket resellers to target customers with products that are being sold in breach of Google's own ad policies and UK law. It is time for companies such as Google to take more responsibility and act against such advertising, or else be considered to be knowingly making money out of fraudulent selling.'*

In this proposal we leave moot the prospect of search being covered by a duty of care but feel that the generic point made by the DCMS committee is a strong one.

Online scams/fraud/identity theft were the second highest reported category of concern (44%) amongst UK adult internet users in survey work by OFCOM and the ICO in Autumn 2018. Fraud and financial loss were the greatest concerns amongst people concerned about these topics and 13% of all adults in the survey have experienced 'Scams / fraud / identity theft'.[159] The Financial Conduct Authority reported an

increase in online contact over telephone contact in attempted scams, rising from 45% in 2017 to 54% in 2018.[160]

Harms to national security

The main harms here comprise violent extremism, terrorism, state sponsored cyber warfare, all of which are illegal. Definitions may be found in the relevant statutes.

Emotional harm

As noted, such harm is unlikely to be covered by the law as it stands. The common law sets a very high bar for emotional harm, of a 'recognised psychiatric injury'[161]. This is out of kilter with the past decade of the law changing – specifically the criminal law (which might be anticipated to require higher standards of harm than civil law given the impact of a guilty finding) - to recognise an emotional component of recognised harms caused by behaviour, often where women are victims – stalking, domestic abuse, harassment[162], controlling or coercive behaviour.[163] We suggest that emotional harm is reasonably foreseeable on some social media and that services should have systems in place to prevent emotional harm suffered by users such that it does not build up to the current threshold of a recognised psychiatric injury. It should be possible to find a form of words that excludes the normal ordinary annoyances of life; Kay LJ used, for example, the formulation 'grievous non-physical reaction'[164]. It may be possible to identify specific examples: eg. harm arising from the abuse of one person by many others or service design that is intentionally addictive. Emotional harm includes harm to vulnerable people – in respect of suicide, self-harm anorexia, mental illness etc (though we note that some of these examples might in any event satisfy the 'recognised psychiatric injury' test).

Harm to young people

Types of harm include bullying, aggression, hate, sexual harassment and communications, exposure to harmful or disturbing content, grooming (recent FoI requests by the NSPCC reveal 5,161 crimes of sexual communications with a child recorded in 18 months[165]), child abuse (See UKCCIS[166] Literature Review) as well as impact on mental health.

Work by 5Rights foundation has described a wide range of harms to children arising from social media and messaging services. The vectors for these harms are often aspects of service design. The statutory duty of care is intended to bite at a systems level, and so would include harmful aspects of design. The duty would cover not just harmful persuasive design, but also careless service design that leads to harm.

5Rights Foundation's January 2019 report "Towards an Internet Safety Strategy"[167] looks at both risk and harms and offers a useful itemised list of harms to children from digital media. See Figure TWO

Harms

A harm is anything that negatively impacts on the health, wellbeing and/or safety of a child. Risks may cause one or a series of harms. They include:

- Loss of confidence
- Isolation
- Sleeplessness
- Over-exposure and over-sharing
- Stress
- Anxiety
- Depression
- Family conflict
- Diminished empathy
- Poor fine motor skills
- Aggression
- Opportunity cost
- Diminished memory, concentration and ability to focus/engage
- Obesity
- Self-harm
- Violent extremism
- Suicide/suicidal ideation
- Misogyny
- Reputational damage
- Loss of autonomy
- Sexual exploitation
- Violence/fear of violence
- Sexual assault
- Addiction/compulsion
- Emotional difficulties/distress
- Behavioural difficulties/distress
- Financial loss
- Permissive and unrealistic sexual attitudes
- Gender-stereotypical sexual attitudes
- Maladaptive attitudes to relationships
- Susceptibility to advertising
- Self-exclusion/self-editing
- Unrealistic body image/pressure to conform to narrow body image
- Discrimination

Primary Sources:

Growing Up With The Internet, Select Committee on Communications [HL], 2017
Mental Health of Children and Young People in England, NHS, 2018
Girls Attitude Survey 2018, Girl Guiding, 2018
Safety Net: Cyberbullying's Impact on Young People's Mental Health, The Children's Society, 2018
Children and Parents: Media Use and Attitudes Report, Ofcom, 2017
Impact of Marketing Through Social Media, Online Games and Mobile Applications on Children's Behaviour, European Commission, 2016
The Datafied Child: The Dataveillance of Children and Implications for Their Rights, New Media and Society, 2017
Young People's Experiences of Online Sexual Harassment, Project deSHAME, 2017
How Technology Hijacks People's Minds, Tristan Harris, 2016
Growing Up Digital Alberta, Harvard Medical School, 2016
Association Between Portable Screen-Based Media Device Access and Sleep Outcomes, JAMA Pediatrics, 2016
Smartphone and Tablet Use: Associations with Sugary Drinks, Sleep, Physical Activity and Obesity, Harvard School of Public Health, 2017
Issue Paper on Youth Radicalisation, Radicalisation Awareness Network, 2018

Figure 2- 5Rights Foundation 'Towards an Internet Safety Strategy'

Harms to justice and democracy

This head of harm is different from the others as, rather than focussing of harms experienced individually, it looks to harms which are impact society as a whole and which tend to have been under-recognised (though the Environmental Protection Act is perhaps an example of such harms being protected by legislation). This point is justified by Emeritus Professor of Philosophy at the University of Cambridge, Baroness O'Neill (speaking in a House of Lords debate) who said:

> *'The harms I have mentioned are all private harms in the economist's sense of the term: they are harms suffered by individuals who are bullied or whose privacy is invaded, or whose education is damaged. There is a second range of less immediately visible harms that arise from digital media. These are public harms that damage public goods, notably cultures and democracy.*[168]

The list of harms includes preventing intimidation of people taking part in the political process beyond robust debate, concerns about disinformation[169] (which may extend beyond discussions of politics[170]), and protecting the criminal and trial process.

Criminal harms

A traditional focus for the debate on internet harms has been the 'if it is illegal offline it should be illegal online' and then to focus on the removal of content that is contrary to the criminal law. While the criminal law may identify types of content that cause significant harm, and would therefore fall within the scope of the regime, the criminal law does not constitute a complete list of harms against which we would expect a service provider to take action. Nor is harm caused only by content but also by the impact of the underlying systems such as software, business processes and their resourcing/effectiveness. We therefore do not think that the question of whether an action constitutes a criminal offence is helpful in determining harms.

Precautionary principle and the definition of relevant harms

In summary, Parliament should set out the primary or heads of harms as a non-exclusive list (as above) to give the regulator initial direction. Using a regulator to provide more detail means that political actors are not involved in questions of what is acceptable at a level of detail. In taking this approach, the regime adopts the approach found in the neighbouring field of broadcasting regulation. In elaborating these harms the regulator must have regard to evidence bearing in mind the precautionary principle (this is different from the obligation on companies in which the steps they take are determined by the principles of reasonableness and forseeability). For example, the UK Chief Medical Officers recent review of the evidence on the impacts of screen time and the mental health and wellbeing of young people, came to the conclusion that:

> *"Even though no causal effect is evident from existing research between screen- based activities, or the amount of time spent using screens, and any particular negative effect, it does not mean that there is no effect. It is still wise to take a precautionary approach. This needs to be balanced, however, against the potential benefits that CYP can derive from their screen-based activities."*[171]

The regulator should make an annual report on trends, research and state of harms in the UK much like OFCOM's market reports and also examine international developments working with other national regulators as part of its responsibility to keep the code up to date.

The Statutory Duty

We would require qualifying social media and messaging service providers to ensure that their service was designed in such a way to be safe to use by reference to the harms (and possible vectors of harms) as identified by the regulator or otherwise foreseeable to the provider, including those arising at a system design level. We expect that the systems operators identify which of those harms are likely to occur on their respective platforms and to take reasonable steps to mitigate that risk. Examples of ways in which risk of harm could be reduced are given in chapter 9. As new harms arise and are identified by the regulator then they are added to the regulatory regime.

The challenge of defining harms

One question relates to the level of detail at which harm must be detailed to be sufficiently clear and, indeed, whether it should be specified in detail in statute in the interests of legitimacy. We have already explained our view that the detail of the harms should be derived from high level statements of relevant harms by the regulator and set down in code. We are not alone in this approach: it was for example discussed by the Digital, Culture, Media and Sport Select Committee in its report on disinformation. [172] We note moreover, that in our experience of regulation, competent regulators have had little difficulty in the past in working out what harm means. In a debate on a statutory duty of care in the House of Lords Baroness Grender made this point:

If in 2003 there was general acceptance relating to content of programmes for television and radio, protecting the public from offensive and harmful material, why have those definitions changed, or what makes them undeliverable now? Why did we understand what we meant by "harm" in 2003 but appear to ask what it is today?[173]

OFCOM's task in the Communications Act 2003 to which Baroness Grender refers is somewhat harder than merely harm:

> *'generally accepted standards are applied to the content of television and radio services so as to provide adequate protection for members of the public from the inclusion in such services of offensive and harmful material'.*[174]

There are other sources which suggest that harms can be identified from broad descriptions these include the AVMSD, OFCOM/ICO research into harms and the Age Appropriate Design Code:

AVMSD - The amendment to the Audio-Visual Media Services Directive was published in November 2018[175]. As we discuss above, it will apply to many social media services that share video. The Directive adapts the concerns found in the traditional audio-visual environment to apply to "video sharing platforms". The Directive – as regards video sharing platforms – specifies that VSPs should be required to "take appropriate measures to protect:

(a) *minors from programmes, user-generated videos and audiovisual commercial communications which may impair their physical, mental or moral development in accordance with Article 6a(1);*

(b) *the general public from programmes, user-generated videos and audiovisual commercial communications containing incitement to violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the Charter;*

(c) *the general public from programmes, user-generated videos and audiovisual commercial communications containing content the dissemination of which constitutes an activity which is a criminal offence under Union law, namely public provocation to commit a terrorist offence as set out in Article 5 of Directive (EU) 2017/541, offences concerning child pornography as set out in Article 5(4) of Directive 2011/93/EU of the European Parliament and of the Council and offences concerning racism and xenophobia as set out in Article 1 of Framework Decision 2008/913/JHA".*

These are reasonably open-textured provisions, that allow a certain room for interpretation by the respective national regulators. Note also that the list of protected characteristics enumerated in Article 21 EU Charter is long:

> *Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.*

In September 2018, OFCOM published with the ICO a joint survey of online harms[176]. This survey is unusual due to its large sample size, professional design and being independent of lobby groups. The survey asked people to gauge the severity of harm. If our proposals are implemented, this could form a very early step towards a harm reduction cycle, which we describe in more detail in chapter 9.

The forthcoming Age Appropriate Design Code[177] will also require a range of risk management measures to be taken at a service design level for services to children, providing another example where a regulator has worked to produce more detailed rules from quite broad Parliamentary requirements. We continue to work closely with 5Rights and the ICO on this matter.

9 How Would a Social Media Harm Regulator Work?

Given the huge power of most social media and messaging service companies relative to an individual we would appoint a regulator to enforce the duty of care; expecting individuals to assert rights through court processes as a mechanism to control the problems on these platforms is unlikely to work given the costs and stresses of litigation, the asymmetry in knowledge and power between the platforms and individual litigants. The regulator would ensure that companies have measurable, transparent, effective systems in place to reduce harm; to do so it would be provided with information -gathering powers[178]. The regulator would have powers of sanction if companies did not comply with their duty of care. In this chapter, we set out the tasks that might be given to a regulator and how the regulator would go about putting them into action.

As noted in chapter 5, the regulator would be independent. The regulator would not get involved in individual items of speech and would of course be bound by the Human Rights Act. In exercising its powers, it must have regard to fundamental human rights and the need to maintain an appropriate balance between conflicting rights. The regulator must not be a censor. Regulators in the UK such as the BBFC, the ASA and OFCOM (and its predecessors) have demonstrated for decades that it is possible to combine quantitative and qualitative analysis of media, neutral of political influence, for regulatory process. Joint OFCOM/ICO research in 2018[179], noted in chapter 8, shows the early stages of quantitative work on harms. Such research is important as it can serve the two-fold role of clarifying the obligations on the operators (an important factor in the context of rights protection) and moving the system towards a reflection of current societal norms.

Risk-based regulation – not treating all qualifying services the same

Central to the duty of care is the idea of risk[180]. We are not proposing that a uniform set of rules apply across very different services and user bases but that the risks and appropriate responses to those risks are assessed in the light of these differences. Harmful behaviours and risk have to be seen in the context of

the platform. In assessing whether a statutory duty of care had been met, the regulator would examine whether a social media service operator has had particular regard to its audience (and in this there are similarities to the assessments made by OFCOM in relation to content regulation). For example, a mass membership, general purpose service open to children and adults should manage risk by setting a very low tolerance for harmful behaviour, in the same way that some public spaces, such as say a family theme park take into account that they should be a reasonably safe space for all. The risk profile would be different for a specialist site targeted at a more limited audience. Specialist audiences/user-bases of social media services may have online behavioural norms that on a family-friendly service could cause harm but in the community where they originate are not harmful. Examples might include sports-team fan services or sexuality-based communities. This can be seen particularly well with Reddit: its user base with diverse interests self-organises into separate subreddits, each with its own behavioural culture and approach to moderation. Mastodon also has distinct communities each of which sets its own community rules (as opposed to ToS imposed by the provider) within the overarching Mastodon framework. In some sites, robust or even aggressive communications (within the law) could be allowed.

One possible benefit of this approach could be that there might be more differentiation between providers and consequently possibly more choice, though we note the strong network effects in this sector. Differentiation between high and low risk services is common in other regulatory regimes, such as for data in the GDPR and is central to health and safety regulation (see chapter 5). In those regimes, high risk services would be subject to closer oversight and tighter rules, as we intend here.

This regime would be implemented in a sector where there is already indicative evidence of harms. A risk management and mitigation process aims to tackle the flow of future harms but there will be an existing challenge of dealing with the stock of harm and the ongoing flow until the regulatory system's risk management objectives are met. We consider here an aspect of the overarching regulatory process, how to reduce manifest harms. This harm reduction process would inform overall risk management and mitigation.

Harm reduction cycle

We envisage an ongoing evidence-based process of harm reduction where the regulator works with the industry and civil society to create a cycle that is transparent, proportionate, measurable and risk-based. The regulator would prioritise high-risk services, and would only have minimal engagement with low-risk services. We describe a cycle here that relies on consultation and feedback loops with the regulator and civil society. However the UK's stylised approach to consultation[181] cycles may be too lengthy for a fast moving industry and the regulator would have to adopt an abbreviated approach.

A harm reduction cycle begins with measurement of harms. Here we emphasise that as the companies' performance is to be managed at system level, we do not intend that the effect of social media and messaging use on each individual should be measured. Rather what is measured is the incidence of artefacts that – according to the code drawn up by the regulator – are deemed as likely to be harmful (to a particular audience) or if novel could reasonably have been be foreseen to cause harm. We use 'artefact' as a catch all term for types of content, aspects of the system (e.g. the way the recommender algorithm works) and any other factors. We discuss foreseeability below.

The regulator would draw up a template for measuring harms, covering scope, quantity and impact. The regulator would then consult publicly on this template, specifically including the qualifying social media

services. The qualifying social media services would then run a measurement of harm based on that template, making reasonable adjustments to adapt it to the circumstances of each service. The regulator would have powers in law to require the companies providing the qualifying services (see enforcement below) to comply. The companies would be required to publish the survey results in a timely manner. This would establish a first baseline of harm.

The companies would then be required to act to reduce these harms by setting out and implementing a harm reduction action plan. We expect those planned actions to be in two groups – things companies just do or stop doing, immediately; and actions that would take more time (for instance new code or terms and conditions changes). Companies should inform the regulator and publish their actions[182]. Companies should seek views on their action plan from users, the victims of harms, the NGOs that speak for users and victims etc. The companies' responses to public comment (though they need not adopt every such suggestion made) would form part of any assessment by the regulator of whether an operator was taking reasonable steps and satisfying its duty of care. Companies would be required to publish their action plans, in a format set out by the regulator, such as:

- what actions they have taken immediately
- actions they plan to take
- an estimated timescale for measurable effect and
- basic forecasts for the impact on the harms revealed in the baseline survey and any others they have identified.

The regulator would take views on the plan from the public, industry, consumers/users and civil society and makes comments on the plan to the company, including comments as to whether the plan was sufficient and/or appropriate. The companies would then continue or begin their harm reduction work.

Harms would be measured again after a sufficient time has passed for harm reduction measures to have taken effect, repeating the initial process. This establishes the first progress baseline.

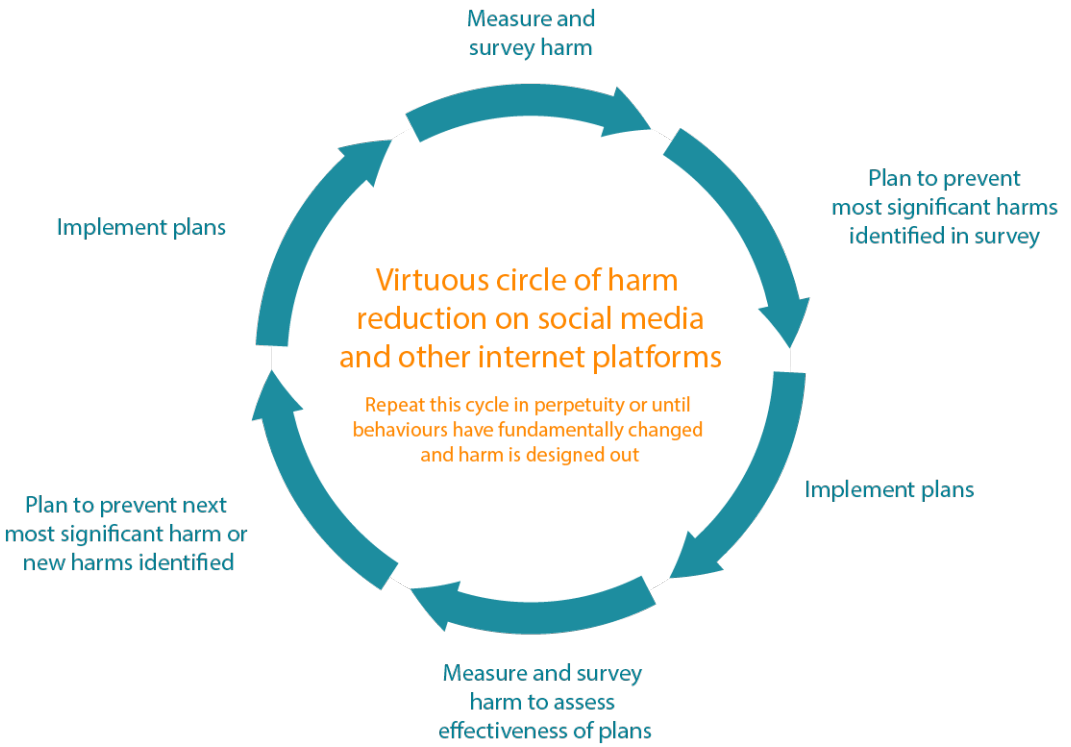
The baseline will reveal four possible outcomes – that harms:

- have risen;
- stayed the same;
- have fallen; or
- new harms have occurred/been revealed.

If harms surveyed in the baseline have risen or stayed the same the companies concerned will be required to act and plan again, taking due account of the views of victims, NGOS and the regulator. In these instances, the regulator may take the view that the duty of care is not being satisfied and, ultimately, may take enforcement action (see below). If incidence of harms has fallen then companies will reinforce this positive downward trajectory in a new plan. Companies would prepare second harm reduction reports/

plans as in the previous round but including learning from the first wave of actions, successful and unsuccessful. Companies would then implement the plans. The regulator would set an interval before the next wave of evaluation and reporting.

Well-run social media services would quickly settle down to a much lower level of harm and shift to less risky service designs. This cycle of harm measurement and reduction would continue to be repeated, as in any risk management process participants would have to maintain constant vigilance of their systems.



We anticipate that well-run services and responsible companies will want to comply with a harm reduction process. Where companies do not comply, or where the regulator has grounds to believe that they have not we propose that the regulator has information gathering powers as is normal in modern regulation (see for example the powers granted to the Information Commissioner). A net output of the harm reduction cycle would be a transparency report produced by each company to ensure an accurate picture of harm reduction was available to the regulator and civic society organisations.

Measurement

While the harm reduction cycle envisages that the use of the platforms by users should be subject to some sort of surveying, we do not envisage that the measurement necessitates constant monitoring of all use and neither statute nor the regulator should require this. At the scale at which many platforms operate, statistical sampling methods should be sufficiently robust combined with other measures. For instance, as we discuss elsewhere the platforms' own mailbox/complaints log should provide early warning of systemic problems.

Foreseeability

Companies who run a qualifying social media or messaging service will have to form judgements about what risks are reasonably foreseeable. This is at the heart of any risk assessment – there will be risks which will be obvious – for instance material harm is known to have occurred before in certain circumstances and those which, while not obvious are foreseeable. If a material risk is foreseeable then a company should take reasonable steps to prevent it. Health and safety law for instance has developed substantial practice around foreseeability of risk which could act as a starting point. Companies will want to ensure that the regulator understands their approach to foreseeability and the regulator will want to discuss that with civil society and victims to help the industry arrive at a consistent interpretation. This is an area where a trade association could help, particularly in providing guidance for SMEs.

From Harms to Codes of Practice

An output of the harm reduction cycle would be industry generated codes of practice that could be endorsed by the regulator. In our view, the speed with which the industry moves would mitigate against traditional statutory codes of practice which require lengthy consultation cycles. The government, in setting up such a regime, should allow some lee-way from standard formalised consultation and response processes. Codes of practice, as well as other forms of guidance (which could also be produced by the regulator) make compliance easier for small companies.

Proportionality

Some commentators have suggested that applying a duty of care to all providers might discourage innovation and reinforce the dominance of existing market players. We do not think that the application of the duty of care would give rise to a significant risk in this regard, for the following reasons.

Good regulators do take account of company size and regulation is applied proportionate to business size or capability[183]. We would expect this to be a factor in determining what measures a company could reasonably have been expected to have taken in mitigating a harm. Clearly, what is reasonable for a large established company would be different for an SME. The 2014 statutory 'Regulators Code' even requires some regulators to take a proportionate, risk managed approach to their work, the code says that:

> *'Regulators should choose proportionate approaches to those they regulate, based on relevant factors including, for example, business size and capacity.'*[184]

The European Commission also acknowledges this in its proposed directive on online terrorist content which requires 'economic capacity' to be taken into account in deciding the adequacy of a company's response.[185]

The proportionality assessment proposed does not just take into account size, but also the nature and severity of the harm, as well as the likelihood of it arising (as discussed below in relation to risk-based regulation). For small start-ups, it would be reasonable for them to focus on obvious high risks, whereas more established companies with greater resources might be expected not only to do more in relation to those risks but to tackle a greater range of harms.

The regulator should determine, with industry and civil society, what is a reasonable way for an SME

service provider to manage risk. Their deliberations might include the balance between managing foreseeable risk and fostering innovation (where we believe the former need not stymie the latter) and ensuring that new trends or emerging harms identified on one platform are taken account of by other companies in a timely fashion. The regulatory emphasis would be on what is a reasonable response to risk, taken at a general level. In this, formal risk assessments constitute part of the harm reduction cycle; the appropriateness of responses should be measured by the regulator against this.

We note that, in other sectors (notably HSWA, data protection and guidance on the Content Codes for broadcasting), regulators give guidance on what is required by the regulatory regime and ways to achieve that standard. This saves businesses the cost of working out how to comply. In addition to guidance as to what risks are likely and immediate steps to mitigate those risks (provided in easier to understand language, perhaps even decision trees), another way to support companies would be the development of libraries of 'good code' that provide appropriate solutions to some of the most common risks[186].

As in other sectors, regulation will create or bolster a market for training and professional development in aspects of compliance. We would expect the regulator to emphasise the need for training for start-ups and SME`s on responsibility for a company's actions, respect for others, risk management etc. The work on ethics in technology could usefully influence this type of training.

Furthermore, regulators would not be likely to apply severe sanctions in the case of a start-up, at least initially. A small company that refused to engage with the regulatory process or demonstrated cavalier behaviour leading to harms would become subject to more severe sanctions. Sanctions are discussed below.

Techniques for harm reduction

We do not envisage that a system that relied purely on user notification of problematic content or behaviour and after the event responses would be taking sufficient steps to reduce harm. In line with the 'by design' approach, consideration for the effects of services and the way they are used should happen from the beginning of the design process and not be mere afterthoughts and consider the way the system as designed affects its users and their behaviour. Nor should the responsibility for safety principally fall to individuals affected (some of whom may not even use the platform) to configure complex tools and to be resilient enough to complain (and persist in that complaint in the fact of corporate disinterest), especially where those people are vulnerable. We draw the following from a wide range of regulatory practice, but the list is not intended to be exhaustive. Some of these the regulator would do, others the regulator would require the companies to do if they were not implementing systems to drive down risk of harm.

If we look to examples of tech companies' response to existing regulatory frameworks, notably data protection, there are some concerns about internal governance and whether issues surrounding harm would be treated with sufficient seriousness and receive the attention of senior management.[187] The UK government has sought to improve the responsibility of Directors and senior staff in the financial services regulatory regime '...in response to the 2008 banking crisis and significant conduct failings such as the manipulation of LIBOR'. The Financial Conduct Authority says[188]:

The aim of the Senior Managers and Certification Regime (SM&CR) is to reduce harm to consumers and strengthen market integrity by making individuals more accountable for their conduct and competence. As part of this, the SM&CR aims to:

- encourage a culture of staff at all levels taking personal responsibility for their actions;
- make sure firms and staff clearly understand and can demonstrate where responsibility lies.

It may be desirable that the statutory duty of care regime sets out some similar structural requirements.

Each qualifying social media service provider could be required to:

- develop a statement of assessed risks of harm, prominently displayed to all users when the regime is introduced and thereafter to new users; and when launching new services or features;
- provide its child protection and parental control approach, including age verification, for the regulator's approval[189];
- develop easy to use tools for users to control the content they see and to limit their exposure to others;
- display a rating of harm agreed with the regulator on the most prominent screen seen by users;
- an internal review system for risk assessment of new services, new tools or significant revision of services prior to their deployment (so that the risk is addressed prior to launch or very risky services do not get launched);
- develop a triage process for emergent problems (the detail of the problem may be unknown, but it is fairly certain that new problems will be arising);
- work with the regulator and civil society on model standards of care in high risk areas such as suicide, self-harm, anorexia, hate crime etc;
- provide adequate complaints handling systems with independently assessed customer satisfaction targets and also produce a twice yearly report on the breakdown of complaints (subject, satisfaction, numbers, handled by humans, handled in automated method etc.) to a standard set by the regulator[190]; and
- Assess how effective the company's enforcement of its own terms of service is and if necessary improve its performance.

The regulator would at a minimum use the following tools and techniques:

- publish model policies on user sanctions for harmful behaviour, sharing research from the companies and independent research;
- detailed guidance as to the meaning of harm;
- publish transparency report models for companies to follow;

- set standards for and monitoring response time to queries (as the European Commission does[191] on extremist content through mystery shopping);
- co-ordinate with the qualifying companies on training and awareness for the companies' staff on harms[192].
- approve industry codes of practice, where appropriate; and
- provide guidance to companies, in particular with SMEs in mind on suggested approaches to dealing with well-known problems and watchlists of issues that might arise when operating particular types of service (or encourage trade bodies to do so);
- monitor if regulated problems move elsewhere and to spread good practice on harm reduction;
- publish a forward-look at non-qualifying social media services brought to the regulator's attention that might qualify in future;
- support research into online harms – both funding its own research and co-ordinating work of others;
- establish a reference/advisory panel to provide external advice to the regulator – the panel might comprise civil society groups, people who have been victims of harm, free speech groups;
- require operators to have in place adequate compliance procedures and sufficient resources to fulfil the regulator's codes and rules[193].

Many commentators call for more media literacy training for children[194]. We think the need for training goes much further. Education is an important tool, not just in developing resilience in users, but also in introducing would-be software developers and service operators to some of the ethical and legal issues. Education could be a mechanism to disseminate that guidance and risk-tested code libraries should be developed and available.[195]

Harm Reduction and the e-Commerce Directive

Although we are not convinced that all qualifying social media companies would be neutral intermediaries, and therefore benefit from intermediary immunity in the e-Commerce Directive (discussed chapter 5), there is a question as whether some of the measures that might be taken as part of a harm reduction plan could mean that the qualifying company loses its immunity, which would be undesirable. There are three comments that should be made here:

- Not all measures that could be taken would have this effect;
- That the Commission has suggested that the e-Commerce Directive be interpreted – in the context of taking down hate speech and other similarly harmful content as not meaning that those which take proactive steps to prevent such content should be regarded as thereby assuming liability;
- After Brexit, there may be more scope for changing the immunity regime – including the change to include an express 'good Samaritan defence'[196].

This harm reduction cycle is similar to techniques used by the European Commission as it works with social media service providers to remove violent extremist content.

Consumer redress

We note the many complaints from individuals that social media services companies do not deal well with complaints – in this the companies may not be unusual. There are many issues with redress mechanisms inside regulatory processes in the UK[197] but this does not mean that matters should be left there. Handling of complaints is an important part of harm reduction – a company's complaints process should function as an early warning of systemic problems. Some aspects of designing for a safe service will require companies to place more emphasis on their redress and complaint mechanisms. Where companies have made a business decision to encrypt and decentralise services, reducing their own ability to detect harm through sampling or software, then they will require a more robust and effective complaint service to investigate and address harms reported to them.[198] As noted, we also envisage that reporting on the operation of the complaints and redress mechanisms would form part of a provider's transparency and reporting obligations. This external review is important not just in terms of effectiveness of harm reduction but in ensuring that the companies are operating fairly, that the meaning of terms of service are clear and consistently applied, that the rights of some groups are not prioritised over those of others and that when take down is in issue, the decision to take content down or not to take it down can be objectively verified.[199]

We have not had the capacity fully to consider ADR/ombudsman regimes inside this regulatory proposal. We are encouraged that doteveryone is leading work on digital redress, and also note the views of Matthew Vickers, CEO of the Ombudsman Services at a recent doteveryone seminar reported in a blog post:

> *'(he) called for us to challenge the paradigm in which redress is a transactional, confrontational and individualised process. Often the person who has sought redress (typically a middle-aged, middle-class, university educated, white male) isn't necessarily left feeling any better after having won their case. They're ultimately left feeling greater distrust of the company they've been dealing with. So how can we help companies realise that good redress equals trust and that trust is a corporate, system and social asset?*

Individual right of action

We do not intend that a new statutory duty of care should remove any rights of an individual to go to court for, say, defamation, nor to complain to the police if that person felt a crime had been committed. But should the statutory duty of care enable an individual right of action to allow someone to sue a company personally for breach of the duty rather than – or in addition to – allowing the regulator to act?

The statutory duty of care is designed to be systemic, preventative and monitored/enforced by a regulator focussing on systemic issues in companies. It may be that individuals would have an interest in these issues rather than the specific outcome in their own cases but, nonetheless, we question whether from the perspective of the provider there is a risk that systemic issues would just be added to claims on substance – a form of double jeopardy. We also note the concerns about a 'compensation culture' raised in relation to HSWA (see chapter 5), as well as concerns in the press about 'ambulance chasing' under data protection rules following a data breach[200]. In our view an individual right of action would create a

complex regulatory environment for companies and the courts. In the general absence of legal aid, facing a highly asymmetric environment would only be available to very few people. For these reasons we think an action for breach of statutory duty would not be appropriate.

There remains the concern that a regulator may have a blind spot and that this could lead to a gap in protection. It may be that other mechanisms involving a designated organisation could be an appropriate safeguard in the event of a dilatory regulator. We suggest that a form[201] of super-complaint mechanism where nominated advocacy groups can bring a complaint to the regulator about aspects of the regulated services that cause harm[202] could be worth considering or a regime such as that found in Article 80 GDPR[203] which allows collective action via mandated groups which can provide compensation to users.

Regulator's interaction with Criminal Law

We set out a series of 'key harms' in chapter 8. Some of these were criminal offences, such as the 'stirring up' offences. Through setting out key harms in the statutory duty of care, we sought to make companies work to mitigate these where they constituted reasonably foreseeable harms. We would envisage mechanisms such as swift and appropriate responses to complaints, consideration of stay-down mechanisms in the context of proven criminal material, and early warning tools for some categories of crime (e.g. patterns of communication in grooming).

These mechanisms would be important in preventing the ongoing commission of crimes (through continuing to distribute criminal content) particularly given that the police have had difficulties coping with the volume of content as well as, in some instances, difficulties understanding the digital environment. This activity would support the eventual outcomes of the work the Law Commission has in train to reforms and clarify a range of communications and harassment offences.[204]

If service providers take action, e.g. through consistent enforcement of its own terms and conditions/ community standards as regards aggressive behaviour or hate speech, this could act as a deterrent to other users – essentially changing the acceptable norms in that particular public space - and as such could be far more effective for victims than waiting for the police to act after the event. The duty of care would not, of course, displace obligations that service providers have in relation to certain criminal content (e.g. in relation to terrorism).

It is however important to remember that there is a prohibition on requiring general monitoring in Article 15 of the e-Commerce Directive, a prohibition that aims to protect both freedom of speech and privacy. These are rights that remain relevant even after Brexit. The extent to which service providers should be obliged to notify content to the authorities or to comply with the authorities (beyond the requirements of the general law) requires further consideration bearing in mind the fundamental rights of all users.

Sanctions and compliance

Some of the qualifying social media services will be amongst the world's biggest companies. In our view the companies will want to take part in an effective harm reduction regime and comply with the law. We note that some large service providers are themselves calling for regulation[205]. The GDPR penalties and sanctions regime (including levelling fines as a proportion of revenue for data breaches, along with the impact of consequent publicity and reputational damage) have yet to be fully exercised by the ICO and may yet provide an effective preventative model. The impact of the CNIL decision against Google

in France will be an early indicator of the effectiveness of the GDPR regime in modifying corporate behaviour.[206]

The companies' duty is to their shareholders. There is an argument that, in order to spend significant shareholder resources on matters for the public good a company management requires regulation. The scale at which these companies operate means that a proportionate sanctions regime is required. We bear in mind the Legal Services Board paper (2014) on Regulatory Sanctions and Appeals processes:

> *'if a regulator has insufficient powers and sanctions it is unlikely to incentivise behavioural change in those who are tempted to breach regulators requirements.'*[207]

As we discussed in chapter 5, the range of mechanisms available within the Health and Safety regime allow the regulator to try improve conditions rather than just punish the operator, so we would propose a similar range of notices (and to some extent the ICO has a similar approach). For those that will not comply, the regulator should be empowered to impose fines (perhaps GDPR magnitude fines if necessary). We have noted in the context of the ICO that a range of investigative powers to support effective enforcement were introduced[208]; we propose that similar powers be given to the regulator here - a comprehensive suite of information gathering powers such as the ICO's ability to make information notices under section 142 of the Data Protection Act 2018 and information orders under section 145.

All regulatory processes leading to the imposition of sanctions should be transparent and subject to a civil standard of proof. The regulator like any public body would be subject to judicial review.

Sanctions would include:

- Administrative fines in line with the parameters established through the Data Protection Act/GDPR regime of up to €20 million, or 4% annual global turnover – whichever is higher. Many types of fines, however, are routinely insured against.

- Enforcement notices – (as used in data protection, health and safety) – in extreme circumstances a notice to a company to stop it doing something. Breach of an enforcement service could lead to substantial fines.

- Enforceable undertakings where the companies agree to do something to reduce harm.

- Adverse publicity orders – the company is required to display a message on its screen most visible to all users detailing its offence. A study on the impact of reputational damage for financial services companies that commit offences in the UK found it to be nine times the impact of the fine.

- Forms of restorative justice – where victims sit down with company directors and tell their stories face to face.

Effective enforcement: Phoenix Companies, Strategic Insolvency and Directors' Disqualification

In discussing sanctions, some consideration needs to be given to the risk of group structures and strategic insolvency being used to lessen the impact of fines, and orders targeted at a particular corporate body. Piercing the corporate veil is a rare occurrence[209] but in some instances statute provides for instances where the veil may be lifted.

The Company Directors Disqualification Act 1986[210] allows for the disqualification of directors for a wide range of offences, including a petition by the Competition and Markets Authority where a person has engaged in types of anti-competitive behaviour, although this latter case is rare. This regime has a function specifically in relation to the problem of phoenix companies. That is, where a company trades and runs into trouble but the persons behind the company avoid financial or regulatory liability by winding the company up and starting again – often to do exactly the same sort of thing. This problem is well illustrated in the data protection sector. SCL Elections, involved in the Cambridge Analytica scandal, has gone insolvent but the parties behind it on the whole still seem to be carrying on business through different corporate vehicles. In a rare example, where a company had not paid a penalty notice imposed by the ICO, the Insolvency Service announced that the director was disqualified because he failed to ensure that the company complied with its statutory obligations.[211]

More generally, the Government amended the Privacy and Electronic Communications (EC Directive) Regulations 2003[212], which deal with direct marketing because the phoenix problem – the changes allow the Information Commissioner the power to fine relevant officers of the companies too where the contravention of the Regulations "took place with the consent or connivance of the officer" or where the contravention is "attributable to any neglect on the part of the officer." Should there be an issue with domestic companies which take a cavalier approach to a statutory duty of care, such an approach may be helpful.

However, it is hard to understand how a disqualification or fine in the UK would bite in relation to a director of a company that was not established under the laws of the United Kingdom. We also note that identifying liable directors, particularly without a licensing regime and where firms may not be registered in the UK, may be problematic.

Effective enforcement: Corporate Criminal Responsibility

In the event that fines plus reputational damage are not considered sufficient deterrence, for example because of the size of the company and the extent of its resources, there are models in which criminal liability is imposed on the company. In fact, it is possible for companies to commit most crimes, but here we consider examples where corporate liability is expressly envisaged. The UK government has explored these approaches since the Fraud Act 2006[213], including – in particular – the Bribery Act 2010[214] . The most recent approach is the Corporate Criminal Offences (CCO)[215] set out in the Criminal Finance Act 2017[216] (building on the Bribery Act) which provides the only defence for a company (against criminal tax evasion) is to show that it has in place adequate procedures to have prevented one of its officers/staff carrying out the offence.

For large companies, this means they have to make a risk assessment and follow a process that sounds much like implementing a statutory duty of care. If the company did not have procedures in place, it would have committed a crime which can result in an unlimited fine and ancillary orders . In our view, this approach, although largely untested, would drive systems-level compliance.

The CCO does not solve the problem of levying an enforceable fine which has sufficient deterrent effect for the largest companies; or even enforcement action uder the Proceeds of Crime Act 2002. However, this approach would result in the company committing a criminal offence which – in addition to public relations and share price concerns – could well have a knock-on effect in other regulatory environments and jurisdictions. For instance, a corporate criminal offence is likely to affect any service that requires a "fit and proper" test.

Effective enforcement: Director's Responsibility

A further approach if stronger sanctions are considered necessary would be to make a director personally liable. Normally, a fine or other sanction would not affect the directors of the company. It is however possible for statute to set out that directors may be subject to an administrative penalty (fine).

In the extreme case of a financial institution failing, senior managers could be charged with a criminal offence, potentially even leading to imprisonment[217]. Use of the criminal law can be seen also in the context of the HSWA 1974. While a penalty of last resort, the criminal penalties are frequently used, with custodial sentences being imposed in 7% of cases involving a conviction and a further 8% of cases resulting in a suspended sentence.[218] Moreover, following a private members bill, the Health and Safety (Offences) Act 2008 was enacted, which not only increased the size of fines but introduced for the first time an option for imprisonment of employees (and not just directors) for health and safety offices where employees may have contributed to a health and safety offence by their consent, connivance or neglect. Custodial sentences would be limited to the worst cases (those that involve 'public outrage'). There are clearly models for extending criminal sanctions to directors and even other employees, but the question is whether – especially given their likely impact on freedom of expression – they are appropriate in this context.

Sanctions for exceptional harm

The scale at which some of the qualifying social media services operate is such that there is the potential for exceptional harm. The more difficult questions relate to what to do in extreme cases. In a hypothetical example – a social media service was exploited to provoke a riot in which people were severely injured or died and widespread economic damage was caused. The regulator had warned about harmful design features in the service, those flaws had gone uncorrected, the instigators or the spreaders of insurrection exploited deliberately or accidentally those features. Or, sexual harm occurs to hundreds of young people due to the repeated failure of a social media company to provide parental controls or age verification in a teen video service. Are fines enough or are more severe sanctions required, as seen elsewhere in regulation?

In extreme cases should there be a power to send a social media services company director to prison or to turn off the service? We have noted that the regulation of health and safety in the UK allows the regulator in extreme circumstances which often involve a death or repeated, sustained breaches to seek a custodial sentence for a director. In the USA the new FOSTA-SESTA package apparently provides for criminal penalties (including we think arrest) for internet companies that facilitate sex trafficking. The introduction of FOSTA-SESTA led swiftly to closure of some dating services and a sex worker forum having its DNS service withdrawn in its entirety.[219] Further, short of sanctions against directors, could a service be turned off? The Digital Economy Act contains power (Section 23) for the age verification regulator to issue a notice to internet service providers to block a website in the UK. While this may appear to be a model for enforcement here, some caution is advisable. It is likely that sites falling within the DEA pornography provisions will mainly carry a similar sort of content; sites such as Twitter which may carry pornography lie outside the regime. Ancillary effects arising from a S23 blocking of a pornography site would therefore be limited compared with the situation in which a large and general platform were blocked – this could, as noted in chapter 4, give rise to concerns about collateral censorship and therefore require really compelling grounds to justify it.

Sanctions and freedom of expression

Throughout discussion of sanctions there is a tension with freedom of speech. The companies are substantial vectors for free speech, although by no means exclusive ones. The state and its actors must take great care not to be seen to be penalising free speech unless the action of that speech infringes the rights of others not to be harmed or to speak. themselves. To some extent, the usual practices of regulators address this concern as enforcement strategies target the worst offenders and the severest penalties are used as a last resort. As we note above, withdrawing the whole service due to harmful behaviour in one corner of it deprives innocent users of their speech on the rest of the platform. However, the scale of social media services mean that acute large scale harm can arise in such services that would be penalised with gaol if such harm had arisen elsewhere in society.

10 Who Should Regulate to Reduce Harm in Social Media Services?

At the outset of this work we described a 'regulatory function'. Our approach was to start with the problem – harm reduction – and work forwards from that, as opposed to starting with a regulator and their existing powers and trying to fit the problem into the shape of their activities. We now address two linked questions:

- why a regulator is necessary, as we have already implied it is; and
- the nature of that regulator.

The Need for a Regulator

The first question is whether a regulator is needed at all if a duty of care is to be created.

Is the fact that individuals may seek redress in relation to this overarching duty (by contrast to an action in relation to an individual piece of content) in the courts not sufficient? At least two pieces of profound legislation based on duties of care do not have 'regulators' as such – the 1957 Occupiers Liability Act[220] and the 1972 Defective Premises Act[221]. By contrast, the 1974 HSWA does rely on a regulator, now the HSE. A regulator can address asymmetries of power between the victim and the harm causer. It is conceivable for a home owner to sue a builder or a person for harm from a building, or a person to sue a local authority for harm at a playground. However, there is a strong power imbalance between an employee and their boss or even between a trade union and a multinational. A fully functioning regulator compensates for these asymmetries. In our opinion there are profound asymmetries between a user of a social media service and the company that runs it, even where the user is a business, and so a regulator is required to compensate for the users' relative weakness.

What Sort of Regulator?

Assuming a regulator is needed, should it be a new regulator from the ground up or an existing regulator upon which the powers and resources are conferred? Need it be a traditional regulator, or would a self or co-regulator suffice? As we shall see below, instances of co-regulation in the communications sector have run into problems. Self-regulation works best when the public interest to be served and those of the industry coincide. This is not the case here.

Whichever model is adopted, the important point is that the regulator be independent (and its members comply with the Nolan Principles[222]). The regulator must be independent not only from government but also from industry, so that it can make decisions based on objective evidence (and not under pressure from other interests) and be viewed as a credible regulator by the public. This is particular important given the fundamental human rights that are in issue in the context of social media. Independence means that the regulator must have sufficient resources, as well as relevant expertise.

A completely new regulator created by statute would take some years before it was operational. OFCOM, for instance, was first proposed in the Communications White Paper in December 2000[223], was created in a paving act of Parliament in 2002 but did not vest and become operational until December 29 2003 at a cost of £120m (2018 prices). In our view harm reduction requires more urgent (and less expensive) action and for this reason we reject the idea, seen in some proposals, that a new sector specific regulator is required.

We therefore propose extending the competence of an existing regulator. This approach has a number of advantages. It spreads the regulator's overheads further, draws upon existing expertise within the regulator (both in terms of process and substantive knowledge) and allows a faster start. We consider that the following (co) regulators should be considered: Advertising Standards Authority (ASA), the British Board of Film Classification (BBFC), the Information Commissioners Office (ICO) the HSE or OFCOM, all of which have long-proven regulatory ability.

The BBFC seems to have its hands full with the age verification regulator from the Digital Economy Act 2017.[224] Notably, the act does not require the age verification regulator to be independent; we have not investigated the extent to which the BBFC would be viewed as independent in terms of its resources and institutional status. Given these points we consider the BBFC would not be best placed to take on such an extensive responsibility as regulating social media. This also raises the question of how well delegated responsibilities work; OFCOM has recently absorbed responsibilities in relation to video on demand, rather than continue to delegate them to ATVOD. While the ASA regulates some content online including material on social media platforms, this is limited to advertisements (including sponsorship and the like). Whether the ASA is as effective in regulating online content as that off-line – a context in which is has long established allies in the distribution chain – is uncertain. The regulation of social media – a reasonably broadly focussed remit – would constitute a change for the ASA: overall it focusses quite tightly on advertising. Adding in the substantial task of grappling with harm social media services more broadly could damage its core functions. The ICO is in 2018-19 at a time of historic change as it assumes a new set of powers and more than doubles in size to achieve similar staff numbers to OFCOM. We judge that in change management terms adding a further new suite of regulatory responsibilities would not be viable. A better route would be for ICO to work closely with the new regulator of the duty of care. The HSE has a strong track record in running a risk-based system to reduce harm in the workplace, including to some extent emotional harm[225]. It has a substantial scientific and research capability, employing over 800 scientists and analysts. However our judgement is that harm reduction in social media service providers require a regulator with deep experience of and specialism in online industries, which is not where the HSE's strengths lie.

Our recommendation is to vest the powers to reduce harm in social media services to OFCOM. OFCOM has over 15 years' experience of digital issues, including regulating harm and protecting young people in broadcasting, a strong research capability, proven independence, a consumer panel, and also resilience in dealing with multinational companies. OFCOM is of a size (£110-£120 annual income and 790 staff)

where, with the correct funding it could support an additional organisational unit to take on this work without unbalancing the organisation. The Commons Science and Technology Select Committee supported this approach and recommended that OFCOM be given responsibility by October 2019[226]. Of course, any regulator responsible for this field would need to work with regulators with responsibilities in contiguous or overlapping areas (e.g. ICO, CMA); OFCOM has some existing experience in so doing.

The normal approach to funding would be to for the larger companies who are regulated to pay for the regulators services and the smaller companies make a nominal contribution. The regulator could also be funded by a small fraction of the revenue planned to be raised by the Treasury from taxing the revenues of internet companies, of which this would be but a tiny percentage. The relative costs of large regulators suggests that the required resource would be in the low tens of millions of pounds.

11 Legislating to implement this regime

Action to reduce harm on social media is urgently needed. We think that there is a relatively quick route to implementation in law. A short bill before parliament would create a duty of care, appoint, fund and give instructions to a regulator.

We have reviewed the very short Acts that set up far more profound duties of care than regulating social media services – The Defective Premises Act 1972 is only seven sections and 28 clauses (very this was a unusually a private members bill written by the Law Commission); the Occupiers Liability Act 1957 is slightly shorter. The central clauses of the HSWA 1974 creating a duty of care and a duty to provide safe machines are brief.

For social media services, a duty of care and key harms are simple to express in law, requiring less than ten clauses or less if the key harms are set out as sub clauses. A duty for safe design would require a couple of clauses. Some further clauses to amend the Communications Act 2003 would appoint OFCOM as the regulator and fund them for this new work. The most clauses might be required for definitions and parameters for the list the regulator has to prepare. We speculate that an overall length of six sections totalling thirty clauses might do it. This would be very small compared to the Communications Act 2003 of 411 Sections, thousands of clauses in the main body of the Act and 19 Schedules of further clauses.

This makes for a short and simple bill in Parliament that could slot into the legislative timetable, even though it is crowded by Brexit legislation. If government did not bring legislation forward a Private Peers/ Members Bill could be considered.

Endnotes

1 Health and Safety at Work Act (1974), available: https://www.legislation.gov.uk/ukpga/1974/37 [accessed 25 March 2019]

2 OECD: The Polluter Pays Principle (1972), available: https://www.oecd-ilibrary.org/environment/the-polluter-pays-principle_9789264044845-en [accessed 25 March 2019].

3 Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L 178/1, available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32000L0031 [accessed 12 March 2019].

4 See e.g. literature cited by Law Commission, Abusive and Offensive Online Communications: A Scoping Report, Law Com No 381 (HC 1682), 1 November 2018, para 2.153

5 United Kingdom Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA), The Precautionary Principle: Policy and Application, available: http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm [accessed 25 March 2019]

6 HM Government, Internet Safety Strategy – Green Paper, October 2017, available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf [accessed 12 March 2019].

7 Ofcom, Internet users' experience of harm online: summary of survey research, 2018, available: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/120852/Internet-harm-research-2018-report.pdf [accessed 12 March 2019].

8 Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1, available: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1552405263176&uri=CELEX:32016R0679 [accessed 12 March 2019].

9 Competition Act, 1998: available: https://www.legislation.gov.uk/ukpga/1998/41/contents

10 Malicious Communications (Social Media) Bill 2016-2017: bill documents available at: https://services.parliament.uk/bills/2016-17/maliciouscommunicationssocialmedia.html

11 Intimidation in Public Life: A Review by the Committee on Standards in Public Life December 2017 Ref: ISBN 978-1-5286-0096-5, Cm 9543, HC 1017273996 2017-18 https://www.gov.uk/government/publications/intimidation-in-public-life-a-review-by-the-committee-on-standards-in-public-life

12 http://understanding.doteveryone.org.uk/

13 Digital Economy Act, 2017: bill documents available at: https://services.parliament.uk/bills/2016-17/digitaleconomy/documents.html [accessed 25 March 2019]

14 Malicious Communications (Social Media) Bill 2016-2017: bill documents available at: https://services.parliament.uk/bills/2016-17/maliciouscommunicationssocialmedia.html

15 Conservative Party, Forward Together: the Conservative Manifesto, May 2017, available at: https://www.conservatives.com/manifesto [accessed 25 March 2019]

16 HM Government, Internet Safety Strategy – Green Paper, October 2017, available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf [accessed 12 March 2019].

17 Prime Minister Speech at Davos, 25 January 2018, available at: https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january [accessed 25 March 2019]

18 Particulars of proposed Designation of Age Verification Regulator, December 2017, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/669567/particulars_of_proposed_designation_of_age-verification_regulator_-_december_2017.pdf [accessed 1 April 2019]

19 William Perrin and Professor Lorna Woods, Harm Reduction in Social Media, blogs and other material available at: https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/ [accessed 25 March 2019]

20 The Communications Select Committee (Lords), "Regulating in a Digital World", March 2019, available at: https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-to-regulate/ [accessed 25 March 2019]

21 HM Government, Internet Safety Strategy – Green Paper, October 2017, available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf [accessed 12 March 2019].

22 Data Protection Act 2018, available at: http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [accessed 25 March 2019]

23 ICO, Call For Views – Age Appropriate Design Code, 2018: available at: https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/call-for-views-age-appropriate-design-code/ [accessed 25 March 2019]

24 Considered by House of Lords Secondary Legislation Scrutiny Committee (Sub-Committee B), 4th Report of Session 2017–19 (HL Paper 218), 8 November 2018, available: https://publications.parliament.uk/pa/ld201719/ldselect/ldseclegb/218/218.pdf (accessed 28th March 2019).

25 Hansard (HL) 11 December 2018 Vol 794, col 1285 et seq; see also House of Lords Secondary Legislation Scrutiny Committee (Sub-Committee B), 4th Report of Session 2017–19 (HL Paper 218), 8 November 2018, available: https://publications.parliament.uk/pa/ld201719/ldselect/ldseclegb/218/218.pdf (accessed 28th March 2019), para 7

26 Professor Lorna Woods, William Perrin and Maeve Walsh, Internet Harm Reduction Proposal, January 2019 available at: https://www.carnegieuktrust.org.uk/publications/internet-harm-reduction/ [accessed 25 March 2019]

27 A duty of care has been recommended by the following organisations to date in 2019: NSPCC (Taming the Wild West Web, February 2019, available at: https://www.nspcc.org.uk/globalassets/documents/news/taming-the-wild-west-web-regulate-social-networks.pdf); Children's Commissioner (Anne Longfield, Children's Commissioner publishes a statutory Duty of Care for online service providers, February 2019, available at: https://www.childrenscommissioner.gov.uk/2019/02/06/childrens-commissioner-publishes-a-statutory-duty-of-care-for-online-service-providers/ [accessed 25 March 2019]); UK Chief Medical Officers (Department of Health and Social Care, UK Chief Medical Officers Commentary on Screen Time and Social Media map of reviews, February 2019, available at: https://www.gov.uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-of-reviews [accessed 25 March 2019]); House of Commons Science and Technology Committee (Impact of Social Media and Screen Use on Young People's Health, January 2019, available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm [accessed 25 March 2019]); UK Labour Party (speech by Tom Watson, January 2019 available at: https://labour.org.uk/press/tom-watson-speech-fixing-distorted-digital-market/ [accessed 25 March 2019]); Lords Communications Committee ("Regulating in a Digital World", March 2019, available at: https://www.parliament.uk/business/committees/committees-a-z/lords-select/communications-committee/inquiries/parliament-2017/the-internet-to-regulate-or-not-to-regulate/ [accessed 25 March 2019]); All-Party Parliamentary Group on Social Media and Young People's Mental Health and Wellbeing/Royal Society of Public Health (#NewFilters, March 2019, available at: https://www.rsph.org.uk/our-work/policy/wellbeing/new-filters.html [accessed 25 March 2019]

28 See our evidence to the Science and Technology Committee (Commons): available at: https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2018/11/29152819/Impact-of-social-media-and-screen-use-on-young-peoples-health1.pdf [accessed 25 March 2019]

29 Rapid release at massive scale August 2017 https://code.fb.com/web/rapid-release-at-massive-scale/

30 Ofcom/ICO, Internet Users Experience of Harms Online: summary of survey research, SEptebmer 2018, available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/120852/Internet-harm-research-2018-report.pdf [accessed 25 March 2019]

31 United Kingdom Interdepartmental Liaison Group on Risk Assessment (UK-ILGRA), The Precautionary Principle: Policy and Application, available: http://www.hse.gov.uk/aboutus/meetings/committees/ilgra/pppa.htm [accessed 25 March 2019]

32 Committee on Standards in Public Life, Intimidation in Public Life: A Review by the Committee on Standards in Public Life, December 2017 (Cm 9543), available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1__2_.pdf [accessed 12 March 2019]

33 Digital, Culture, Media and Sport Select Committee, Disinformation and 'fake news': Final Report, Eighth Report of Session 2017-19 (HC 1791)

34 See Lawrence Lessig, "The Law of the Horse: What Cyberlaw Might Teach", (1999), 113 Harv. L. Rev. 501; also "Code and Other Laws of Cyberspace" (Cambridge MA: Basic Books, 1999) and "Code: version 2.0" (Cambridge MA: Basic Books,2006)

35 A. Murray, The Regulation of Cyberspace: Control in the Online Environment (Abingdon: Routledge-Cavendish, 2007)

36 C. R Sunstein, 'Nudging: A Very Short Guide' (2014) 37 Journal of Consumer Policy 583

37 For example see the impact of design in relation to online contracts in B. Frischmann and E. Selinger Re-engineering Humanity (Cambridge: Cambridge University Press, 2018), pp. 74-77

38 Written evidence from Medconfidential to DCMS Committee Inquiry on Immersive and Addictive Technologies January 2019 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/immersive-and-addictive-technologies/written/95400.html

39 W. Hartzog, Privacy's Blueprint: The Battle to Control the Design of New Technologies (Cambridge, MA: Harvard University Press, 2018)

40 Article: 'Sean Parker unloads on Facebook: "God only knows what it's doing to our children's brains"' Axios, Mike Allen Nov 9, 2017 https://www.axios.com/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792-f855e7b4-4e99-4d60-8d51-2775559c2671.html; more recent journalism suggests that YouTube also ignored risks to users or to the information environment in the search for engagement: M. Bergen, 'YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant' Bloombeg, 2 April 2019, available: https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant (accessed 3 April 2019).

41 See W Hartzog, Privacy's Blueprint: The Battle to Control the Design of New Technologies (Cambridge, MA: Harvard University Press, 2018), p. 179

42 On difficulties with 'baked in' assumptions see e.g. Antignac, Th, and Le Métayer D., 'Privacy by design: From technologies to architectures' in Preneel B. and Ikonomou D. (eds.) Privacy Technologies and Policy, Springer International Publishing, 2014, pp. 1-17; Koops, B-J., and Leenes, R., 'Privacy Regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data protection law' (2014) 28 International Review of Law, Computers and Technology 159.

43 Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L178/1.

44 Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) (codified version) [2010] OJ L95/1 as amended by Directive 2018/1808 [2018] OJ l 303/69, unofficial consolidated version available: http://data.europa.eu/eli/dir/2010/13/2018-12-18.

45 L. Woods, "Article 11- Freedom of Expression and Information" in Peers, Hervey, Kenner and Ward (eds) The EU Charter of Fundamental Rights: A Commentary (Oxford: Hart Publishing, 2014)

46 Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 Decl (13/02/2019)1 https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

47 Smajić v Bosnia and Herzegovina (App no 48657/16), decision 16 January 2018, para 33.

48 Eon v France, (App no. 26118/10), decision 14 March 2013.

49 Lingens v Austria (App no 9815/82) (1986) 8 EHRR 407, [1986] ECHR 7, para 42; Castells v. Spain (App No 11798/85), A/236, (1992) 14 EHRR 445

50 Steel and Morris v UK (App no 68416/01) [2005] ECHR 103, (2005) 41 EHRR 403para 94

51 Delfi AS v Estonia [GC] (App no 64569/09) [2015] ECHR 586.

52 For approach to this question see e.g Cump n and Maz re v. Romania (App no 33348/96), ECHR 2004-XI, para 90

53 Tolstoy Miloslavsky v. the United Kingdom (App no 18139/91), (1995) 20 EHRR 442, para 35 and 49; c.f. Independent News and Media and Independent Newspapers Ireland Limited v. Ireland (App no 55120/00), (2006) 42 EHRR 1024

54 Observer and Guardian v. United Kingdom (Spycatcher) (App no 13585/88), [1991] 14 EHRR 153, [1991] ECHR 49

55 Yildirm v Turkey (App no 3111/10) ECHR 2012-VI.

56 Article 17 ECHR provides:
Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention.

57 Belkacem v. Belgium (App no 34367/14), decision 27th June 2017.

58 Ibid, para 33; see also M'Bala M'Bala v France (App no 25239/13) decision 20th October 2015 (in re Holocaust denial).

59 See e.g. A. Buyse, 'Dangerous Expressions: the ECHR, Violence and Free Speech' (2014) 63 ICLQ 491.

60 López Ostra v. Spain (App No) 16798/90, judgment 9th December 1994, para 51

61 See generally e.g J. Waldron The Harm in Hate Speech (Cambridge, MA: Harvard University Press, 2012)

62 UN Special Rapporteur on Violence against Women, Report on online violence against women and girls from a human rights perspective (A/HRC/38/47), 14 June 2018, available: https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session38/Documents/A_HRC_38_47_EN.docx (accessed 28 March 2019).

63 Manole v Moldova (App no 13936/02), judgment 17 September 2009, which concerned the obligations of the state monopoly brioadcaster to transmit impartial, independent and balanced news, information and comment, and provide a forum for public discussion in which as broad a spectrum as possible of views and opinions can be expressed.

64 We assume that where platforms take action in response to requirements from the State (e.g. being required to take down child pornography), they effectively act as State agents and that the proper framework for analysis is that of the negative obligation. For cases where the actions of private actors have been imputed to the State, see e.g. Vgt Verein Gegen Tierfabriken v Switzerland (App no. 24699/94), judgment 28 Jun 2001, (2002) 34 EHRR 159, [2001] ECHR 412; c.f. Animal Defenders v UK (App no 48876/08), judgment 22 April 2013. Note this is a different context than cases where an individual seeks a broadcasting licence.

65 A further question which lies outside the immediate scope of this report is the extent to which platform design might be said to affect freedom of expression, either in terms of compelled speech or prohibited speech.

66 Paul de Hert and Dariusz Kloza 'Internet (Access) as a New Fundamental Right: Inflating the Current Rights Framework?' (2012) 3 EJLT available: http://ejlt.org/article/view/123/268

67 Murphy v Ireland (App no. 44179/98), judgment 10 July 2003, [2003] ECHR 352, (2003) 38 EHRR 212.

68 Saliyev v Russia (App no 35016/03) [2010] ECHR 1580.

69 Appleby v United Kingdom (App no 44306/98) ECHR 2003-VI, paras [42]–[43] and [47]–[49].

70 Yildirm v Turkey (App no 3111/10) ECHR 2012-VI.

71 P. de Hert and D. Kloza, ''Internet (access) as a new fundamental right. Inflating the current rights framework?' (2012) 3(3) European Journal of Law and Technology, online publication, available: http://ejlt.org//article/view/123/268; see Jankovis v Lithuania (App no 21575/ 08), decision 17 January 2017 (access to the internet based on right to educatoin) and, by analogy, Khurshid Mustafa v Sweden (App no 23883/06) [2008] ECHR 1710.

72 See e.g. M.L. and W.W. v. Germany (App no 60798/10 and 65599/10) judgment 28 June 2018, [2018] ECHR 554 (right to be forgotten).

73 Pfeifer v Austria (App no 12556/03) [2007] ECHR 935, para 33.

74 See L. Woods 'Social Media: it is not just about Article 10' in D. Mangan & L. Gillies (eds) The Legal Challenges of Social Media (Cheltenham: Edward Elgar, 2017)

75 KU v Finland (App no 2872/02) ECHR 2008-V; Aksu v Turkey (App nos 4149/04 and 4102/04) [GC] ECHR 2012-I; Đorđević v. Croatia, (App no 41526/10), judgment 24 July 2012

76 KU v Finland App No 2872/02 ECHR 2008-V

77 See L. Woods 'Social Media: It is not just about Article 10' in D. Mangan and L. Gillies (eds) The Legal Challenges of Social Media (Cheltenham: Edward Elgar, 2017), p. 121 et seq.

78 UN Office of the High Commissioner for Human Rights, Press Release: UN experts urge States and companies to address online gender-based abuse but warn against censorship, 8 March 2017, available: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21317&LangID=E (accessed 28 March 2019).

79 Article 17(e) UN Convention on the Rights of the Child, available: https://www.unicef.org.uk/what-we-do/un-convention-child-rights/

80 Article 4 e-Commerce Directive.

81 Recital 40 e-Commerce Directive.

82 Joined Cases C-236-8/08 Google [201] ECR I-2417 and Case C-324/09 L'Oreal v eBay [2011] ECR I-6011.

83 Case C-324/09 L'Oreal v eBay, para 124 and see para 122 for examples of activity that a diligent economic operator may engage in; the test of 'diligent economic operator' was applied by the Northern Irish Court of Appeal in C.G. v Facebook Ireland Ltd [2016] NICA 54, para 72.

84 Directive 2004/48/EC on the enforcement of intellectual property rights [2004] OJ L 157/1.

85 UN Office of the High Commissioner for Human Rights, Press Release: UN experts urge

States and companies to address online gender-based abuse but warn against censorship, 8 March 2017, available: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21317&LangID=E (accessed 28 March 2019).

86 Commission, Proposal for a Directive on Copyright in the Digital Single Market (COM (2016) 593 final) 14 September 2016; the text of the directive has been agreed in trilogue but not yet formally adopted.

87 Commission, Proposal for a Regulation on Preventing the Dissemination of Terrorist Content Online (COM (2018) 640 final), 12 September 2018; the text of the directive has been agreed in trilogue but not yet formally adopted.

88 For further discussion see L. Woods, 'Video-sharing platforms in the revised Audiovisual Media Services Directive' (2019) 23 Comms Law 127.

89 See: I. Walden, 'Telecommunications Law and Regulation: An Introduction' in I. Walden (ed) Telecommunications Law and Regulation (5th ed) (Oxford: Oxford University Press, 2018)

90 Communications Act, 2003 available at: https://www.legislation.gov.uk/ukpga/2003/21/contents [accessed 25 March 2019]

91 Directive 2002/20/EC on the authorisation of electronic communications networks and services (Authorisation Directive) [2002] OJ L 108/21, available: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32002L0020 [accessed 13 March 2019].

92 For more detail see e.g. A. Flanagan, 'Authorization and Licensing' in I. Walden (ed) Telecommunications Law and Regulation (5th ed) (Oxford: Oxford University Press, 2018)

93 Ofcom's General Conditions, January 2016, available at: https://www.ofcom.org.uk/advice-for-businesses/knowing-your-rights/gen-conditions [accessed 25 March 2019]

94 Ofcom, Ofcom approved code of practice for customer service and complaints handling: annex to General Condition C4, available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0025/132829/Ofcom-approved-complaints-code-of-practice.pdf [accessed 25 March 2019]

95 Digital Economy Act, 2017, available at: http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted [accessed 25 March 2019]

96 Digital Economy Act, Part 3, 2017, available at: http://www.legislation.gov.uk/ukpga/2017/30/part/3/enacted [accessed 25 March 2019]

97 HM Government, Government response to the Internet Safety Strategy Green Paper, May 2018, p. 23, available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf [accessed 13 March 2019].

98 BBFC, Guidance on Age Verification Arrangements, October 2018, available: https://www.ageverificationregulator.com/assets/bbfc-guidance-on-age-verification-arrangements-october-2018-v2.pdf [accessed 12 March 2019].

99 BBFC, Guidance on Ancillary Service Providers, October 2018, available: https://www.ageverificationregulator.com/assets/bbfc_guidance_on_ancillary_service_providers_october_2018-v2.pdf [accessed 12 march 2019].

100 Explanatory Memorandum to the Draft Online Pornography (Commercial Basis) Regulations 2018, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/749750/Explanatory_Memorandum_to_the_Draft_Online_Pornography__Commercial_Basis__Regulations_2018.pdf [accessed 12 March 2019]

101 Data Protection Act 2018, available at: http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [accessed 12 march 2019]

102 Section 123 Data protection Act 2018, available: http://www.legislation.gov.uk/ukpga/2018/12/section/123/enacted [accessed 12 March 2019].

103 Baroness Kidron, Data Protection Bill, Hansard, 11 December 2017, Vol 787, col 1427, available: https://hansard.parliament.uk/lords/2017-12-11/debates/154E7186-2803-46F1-BE15-36387D09B1C3/DataProtectionBill(HL) [accessed 12 March 2019].
104 ICO, Regulatory Action Policy, available: https://ico.org.uk/media/about-the-ico/documents/2259467/regulatory-action-policy.pdf [accessed 21 March 2019].
105 S. 149 Data Protection Act 2018.
106 Now s 142 Data Protection Act 2018.
107 s. 146 Data Protection Act 2018
108 See e.g. Which's evidence to the Data protection Bill Committee, available: https://publications.parliament.uk/pa/cm201719/cmpublic/dataprotection/memo/dpb32.pdf [accessed 13 March 2019].
109 Health and Safety at Work Act, 1974, available: https://www.legislation.gov.uk/ukpga/1974/37 [accessed 25 March 2019]
110 The Control of Major Hazards Regulations, 1999, available: http://www.legislation.gov.uk/uksi/1999/743/contents/made [accessed 25 March 2019]
111 R. E. Löfstedt Reclaiming health and safety for all: An independent review of health and safety legislation, November 2011 (Cm 8219), available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/66790/lofstedt-report.pdf [accessed 13 March 2019], para 5.
112 Environmental Protection Act, 1990, available: https://www.legislation.gov.uk/ukpga/1990/43/contents [accessed 25 March 2019]
113 The Waste (England and Wales) regulations 2011, available: http://www.legislation.gov.uk/uksi/2011/988/contents/made [accessed 25 March 2019]
114 European Commission, Directive 2008/98/EC on waste (Waste Framework Directive), available: http://ec.europa.eu/environment/waste/framework/index.htm [accessed 25 March 2019]
115 See: M. Bunting, Keeping Consumers Safe Online: Legislating for platform accountability for online content (Communications Cambers), July 2018, p. 20
116 Donoghue v. Stevenson [1932] AC 562 (HL).
117 Gregg v. Scott [2005] UKHL 2, at para 217.
118 Caparo Industries plc v Dickman [1990] 2 AC 605 (HL); but 2018 saw a slew of cases on the scope of duty of care which rinter alia, estricts Caparo to new categories of liability, raising the question of what a new category is– Robinson v. Chief Constable of the West Yorkshire Police [2018] UKSC 4; Steel v NRAM Ltd [2018] UKSC 13; James-Bowen & Ors v Commissioner of the Police of the Metropolis [2018] UKSC 40; Darnley v Croyden Health Services NHS Trust [2018] UKSC 50.
119 Viscount Kilmuir: 'justice has been achieved, to quote the Law Reform Committee, "in spite of, rather than with the assistance of, the categories, which tend to embarrass justice by requiring what is essentially a question of fact to be determined by referring to an artificial and irrelevant rule of law."' HL Deb 21 June 1956 vol 197 cc1181-96 https://api.parliament.uk/historic-hansard/lords/1956/jun/21/occupiers-liability-bill-hl
120 Law Commission, Liability for Psychiatric Illness (LC249), 10 March 1998, available: https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2015/04/LC249.pdf [accessed 12 March 2019].
121 Everett v Comojo [2011] EWCA Civ 13, available: http://www.bailii.org/ew/cases/EWCA/Civ/2011/13.html [accessed 12 March 2019], para 33.
122 Heyes v Pilkington Glass Ltd [1998] PIQR P303,(CA)
123 NSPCC Taming the Wild West Web How to regulate social networks and keep children safe from

abuse February 2019 https://www.nspcc.org.uk/globalassets/documents/news/taming-the-wild-west-web-regulate-social-networks.pdf

124 Twitch corporate blog announcing changes (8 February 2018) https://blog.twitch.tv/twitch-community-guidelines-updates-f2e82d87ae58 [accessed 25 March 2019]: 'We may take action against persons for hateful conduct or harassment that occurs off Twitch services that is directed at Twitch users.' (Twitch Community Guidelines on Harassment: https://www.twitch.tv/p/legal/community-guidelines/harassment/) [accessed 25 March 2019]

125 As the service providers do to counter terrorism in the Global Internet Forum to Counter Terrorism https://gifct.org/about/

126 Health and Safety at Work Act, 1974, available: https://www.legislation.gov.uk/ukpga/1974/37 [accessed 25 March 2019]

127 M. Moore Tech Giants and Civic Power (2016), p. 5

128 M. Bunting, Keeping Consumers Safe Online: Legislating for platform accountability for online content (Communications Cambers), July 2018, p. 22. The Law Commission Report contains a glossary of terms in which it defines terms including 'social media platform' and 'social networking service', see Law Commission, Abusive and Offensive Online Communications: A Scoping Report (Law Com 381) (HC 1682), 1 November 2018,p. ix and p. 2.31

129 https://www.asa.org.uk/codes-and-rulings/advertising-codes.html

130 Unlocking digital competition, Report of the Digital Competition Expert Panel (Furman et al, HMG March 2019) https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel and also letter from the Chair of the Competition and Markets Authority Andrew Tyrie to BEIS 'A letter and summary outlining proposals for reform of the competition and consumer protection regimes from the Chair of the Competition and Markets Authority' February 2019 https://www.gov.uk/government/publications/letter-from-andrew-tyrie-to-the-secretary-of-state-for-business-energy-and-industrial-strategy

131 See for instance Adi Robertson, 'A misogynist Twitch rant has streamers calling for clearer rules', The Verge, November 2017, available: https://www.theverge.com/2017/11/16/16654800/twitch-gaming-trainwrecks-banned-female-streamers-rant-moderation [accessed 25 March 2019]

132 Julia Alexander 'As Discord nears 100 million users, safety concerns are heard' 7 December 2017 https://www.polygon.com/2017/12/7/16739644/discord-100-million-users-safety

133 Angus Crawford, 'Kik chat app 'involved in 1,100 child abuse cases', BBC News, 21 September 2018, available at https://www.bbc.co.uk/news/uk-45568276 [accessed 25 March 2019]

134 Josh Constine, 'Whatsapp has an encrypted child porn problem', Tech Crunch, 20 December 2018, available: https://techcrunch.com/2018/12/20/whatsapp-pornography/ [accessed 25 March 2019

135

136 Sample maximum group sizes: FB Messenger – 150; WhatsApp - 256 (although through tweaking it might be possible exceed this); Snap – 15; iMessage – 20; Kik – 50; Zoom – 2000; Bubble seems to be unlimited; Telegram has grown its group size swiftly to 100,000.

137 Zeynep Tufecki, 'You Tube, the Great Radicalizer', New York Times, 10 March 2018, available: https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html?smid=pl-share [accessed 25 March 2019]

138 Anthropological research suggests that those coding recommender algorithms see their function as 'hooking' users; that these algorithms operate as a trap: N. Seaver, 'Captivating algorithms: Recommender systems as traps' (2018) Journal of Material Culture, available: https://journals.sagepub.com/doi/10.1177/1359183518820366 [accessed 25 March 2019]

139 'Google removed "Are Jews evil?" from its auto-complete function in December 2016 (following a series of articles on the Guardian/Observer website)', See: Community Security Trust, 'What Google Searches Tell Us About Antisemitism today', January 2019, available: https://cst.org.uk/news/blog/2019/01/11/hidden-hate-what-google-searches-tell-us-about-antisemitism-today [accessed 25 March 2019]

140 Josh Constine, 'Microsoft Bing not only shows child pornography, it suggests it', January 2019, available at: https://techcrunch.com/2019/01/10/unsafe-search/ [accessed 25 March 2019]

141 Such as due prominence: Ofcom, EPG Prominence: a report on the discoverability of PSB and local TV Services, July 2018, available: https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/epg-prominence [accessed 25 March 2019]

142 See generally e.g. J. von Hoboken, Search Engine Freedom: On the Implications of the Right to Freedom of Expression for the Legal Governance of Web Search Engines (Alphen a/d Rijn: Kluwer Law International, 2012)

143 HMG response to Internet Safety Green Paper May 2018 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf [accessed 25 March 2019]

144 https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/call-for-views-age-appropriate-design-code/

145 Facebook limits access to its APIs following Cambridge Analytica scandal, see Facebook newsroom, 4 April 2018, available at: https://newsroom.fb.com/news/2018/04/restricting-data-access/ [accessed 25 March 2019]

146 Lords Select Committee report on AI, AI in the UK: ready, willing and able?, April 2018, available at: https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/news-parliament-2017/ai-report-published/ [accessed 25 March 2019]

147 Food Standards Agency, Food Hygiene for your Business, available at: https://www.food.gov.uk/business-guidance/food-hygiene-for-your-business-0 [accessed 25 March 2019]

148 On the advantages of a code see e.g. M. Bunting, Keeping Consumers Safe Online: Legislating for platform accountability for online content (Communications Cambers), July 2018, p. 22.

149 HMG response to Internet Safety Green Paper May 2018, Annex B – Draft code of practice for providers of online social media platforms May 2018 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/708873/Government_Response_to_the_Internet_Safety_Strategy_Green_Paper_-_Final.pdf [accessed 25 March 2019]

150 Review announced following campaign by Stella Creasy MP. See: https://www.bbc.co.uk/news/uk-politics-45423789 [accessed 25 March 2019]

151 House of Commons Petitions Committee Online abuse and the experience of disabled people First Report of Session 2017–19: 'If the criminal law cannot deal with distributing fake child pornography to mock a disabled child and his family, then the law is inadequate. We agree with the petitioner, Katie Price, and the Law Commission that the current law on online abuse is not fit for purpose.'

152 National Trading Standards eCrime Team Alerts (March 2019) http://www.tradingstandardsecrime.org.uk/alerts/

153 Measuring the Cost of Cybercrime – Anderson et al WEIS Conference Paper · January 2012 https://www.econinfosec.org/archive/weis2012/program.html

154 Public Accounts Committee 'The growing threat of online fraud' December 2017 https://publications.parliament.uk/pa/cm201719/cmselect/cmpubacc/399/39903.htm#_idTextAnchor004

155 Caroline Normand, Director of Policy, Which? – supplementary written evidence to Lords Communications Committee inquiry on Internet Regulation (IRN0123) http://data.parliament.

uk/writtenevidence/committeeevidence.svc/evidencedocument/communications-committee/the-internet-to-regulate-or-not-to-regulate/written/92938.html)

156 Which? "Control, Alt or Delete? : the future of consumer data", July 2018, available at: https://www.which.co.uk/policy/digitisation/2659/control-alt-or-delete-the-future-of-consumer-data-main-report)

157 Trading standards successes IP crime and enforcement report 2017 to 2018 https://www.gov.uk/government/publications/annual-ip-crime-and-enforcement-report-2017-to-2018

158 House of Commons Digital, Culture, Media and Sport Committee Live music Ninth Report of Session 2017 https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/733/733.pdf

159 OFCOM/ICO Internet users' experience of harm online 18 September 2018 https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/internet-use-and-attitudes/internet-users-experience-of-harm-online

160 New FCA warning about investment scams – Getsafeonline February 6th 2019 https://www.getsafeonline.org/news/new-fca-warning-about-investment-scams/

161 In relation to tort, see e.g. E. Descheemaeker, 'Rationalising Recovery for Emotional Harm in Tort Law' (2018) LQR 602; R. Mulheron 'Rewriting the Requirement for a "recognised Psychiatric Injury" in Negligence Claims' (2012) 33 OJLS 1.

162 The CPS guidance provides a good overview of harassment issues. See: https://www.cps.gov.uk/legal-guidance/stalking-and-harassment [accessed 25 March 2019]

163 The CPS guidance is a good overview. See: https://www.cps.gov.uk/legal-guidance/controlling-or-coercive-behaviour-intimate-or-family-relationship [accessed 25 March 2019]

164 Hussain v CC of West Mercia Constabulary [2008] EWCA Civ 1205

165 NSPCC 'Over 5,000 online grooming offences recorded in 18 months' News 1 March 2019, available: https://www.nspcc.org.uk/what-we-do/news-opinion/over-5000-grooming-offences-recorded-18-months/ (accessed 3 April 2019)

166 Sonia Livingstone, et al, Children's online activities, risks and safety: a literature review by the UKCCIS Evidence Group, October 2017, available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650933/Literature_Review_Final_October_2017.pdf [accessed 25 March 2019]

167 Baroness Kidron, et al, Disrupted Childhood: the cost of persuasive design, June 2018, available: https://5rightsfoundation.com/static/5Rights-Disrupted-Childhood.pdf [accessed 25 March 2019]

168 (Baroness) Onora O'Neill is Emeritus Professor of Philosophy at the University of Cambridge. Debate 17 January 2019: https://hansard.parliament.uk/ Lords/2019-01-17/debates/3D73C90D-4375-4494-9B17-D6A5A0ED9389/ChildrenAndYoungPeopleDigitalTechnology#contribution-7D902E43-2B67-42F9- 8EC9-AAD1F6F5313B

169 See Digital, Culture, Media and Sport Select Committee, Disinformation and 'fake news': Final Report, Eighth Report of Session 2017-19 (HC 1791), 18th February 2019; LSE Commission on Truth Trust and Technology Tackling the Information Crisis: A Policy Framework for Media System Resilience, chapter 2

170 Royal Society for Public Health, Moving the Needle: Promoting vaccination uptake across the life course, 2018, pp. 29 and 31

171 Department of Health and Social Care, UK Chief Medical Officers Commentary on Screen Time and Social Media map of reviews, February 2019, available at: https://www.gov.uk/government/publications/uk-cmo-commentary-on-screen-time-and-social-media-map-of-reviews [accessed 25 March 2019]

172 Digital, Culture, Media and Sport Select Committee, Disinformation and 'fake news': Final Report, Eighth Report of Session 2017-19 (HC 1791), 18th February 2019, paras 31-32.

173 House of Lords Hansard, Social Media Services, 12 November 2018, vol 793, available : https://hansard.parliament.uk/Lords/2018-11-12/debates/DF630121-FFEF-49D5-B812-3ABBE43371FA/SocialMediaServices [accessed 25 March 2019]

174 Indeed, Baroness Grender refers to only one of Ofcom's duties set out in S319 Communications Act 2003, see: https://www.legislation.gov.uk/ukpga/2003/21/section/319 [accessed 25 March 2019]

175 Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32010L0013 [accessed 25 March 2019]

176 Ofcom, Internet users' experience of harm online: summary of survey research, 2018, available: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/120852/Internet-harm-research-2018-report.pdf [accessed 12 March 2019].

177 ICO, Age Appropriate Design Code – responses to consultation, available: https://ico.org.uk/about-the-ico/responses-to-the-call-for-evidence-on-the-age-appropriate-design-code/ [accessed 25 March 2019]

178 On the importance of evidence gathering powers, see the evidence of Sharon White to the DCMS Select Committee on Disinformation and 'fake news': Final Report, Eighth Report of Session 2017-19 (HC 1791), 18 February 2019, para 33

179 Internet users' experience of harm online 18 September 2018 https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/internet-use-and-attitudes/internet-users-experience-of-harm-online

180 The HSE guidance on risk assessment demonstrates many approaches for companies. See: http://www.hse.gov.uk/risk/ [accessed 25 March 2019]
Also recent HMG guidance on risk and public bodies demonstrates an approach for board members in large organisations. See HMG, Management of Risk in Government, January 2017, available: https://www.gov.uk/government/publications/management-of-risk-in-government-framework HMT guidance on risk management gives a good overview of top down risk management practices. See HMT, Orange Book: Management of Risk – principles and concepts, 2013, available: https://www.gov.uk/government/publications/orange-book [all accessed 25 March 2019]

181 Public bodies often follow the consultation principles set out by the Cabinet Office here: https://www.gov.uk/government/publications/consultation-principles-guidance

182 Google for instance published a blog post explaining why it had suspended commenting on videos with a potential for grooming: "More updates on our actions related to the safety of minors on YouTube", February 2019, available: https://youtube-creators.googleblog.com/2019/02/more-updates-on-our-actions-related-to.html [accessed 25 March 2019]

183 HSE, Health and Safety Made Simple: 'For many businesses, all that's required is a basic series of practical tasks that protect people from harm and at the same time protect the future success and growth of your business.' Available: http://www.hse.gov.uk/simple-health-safety/ [accessed 25 March 2019]

184 HMG, Regulators Code, April 2014, available: https://www.gov.uk/government/publications/regulators-code

185 European Council Press Release, 'Terrorist content online: Council adopts a new negotiating position on new rules to prevent dissemination', 6 December 2018, available: https://www.consilium.europa.eu/en/press/press-releases/2018/12/06/terrorist-content-online-council-adopts-negotiating-position-on-new-rules-to-prevent-dissemination/ [accessed 25 March 2019]

186 For instance Google, Microsoft, Facebook have long worked in hashing of child abuse images, now brokered by IWF (see: https://www.iwf.org.uk/news/tech-breakthrough-announced-on-20th-anniversary-of-iwfs-first-child-sexual-abuse-imagery). Similar action occurs on terrorism: see: https://www.blog.google/outreach-initiatives/public-policy/stop-terror-content-online-tech-companies-need-work-together/ Other developers also share code on harm reduction – for instance this abuse detection code on GitHub (we have not tested the code at link) https://github.com/topics/abuse-detection [all accessed 25 March 2019]

187 See DCMS Committee, Disinformation and 'fake news': Final report, Eighth Report of Sesssion 2017-19 (HC 1791), 18 February 2019, paras 62-63

188 Financial Conduct Authority web article: Senior Managers and Certification Regime First published: 07/07/2015 Last updated: 07/02/2019 https://www.fca.org.uk/firms/senior-managers-certification-regime

189 This would have to mesh with the approach for implementing the Age Appropriate Design Code and the Audio Visual Media Services Directive.

190 The need for an adequate complaints handling mechanism can be found in the telecommunications regime (see chapter 5); the House of Lords Communications Committee noted the need for consistent enforcement as well as transparency of complaints handling: Growing up with the Internet 2nd Report of Session 2016–17 (HL Paper 130), 21 March 2017, paras 241-2

191 Results of Commission's last round of monitoring of the Code of Conduct against online hate speech
19/01/2018 https://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=612086

192 See examples in health and safety regulation, eg HSE Health & Safety Training Courses available: https://www.hsl.gov.uk/health-and-safety-training-courses/ [accessed 25 March 2019]

193 OFCOM radio licence condition requiring 'putting in place adequate compliance procedures to ensure that the licensee can comply with its licence conditions and Ofcom's codes and rules' OFCOM Apply for a radio broadcast licence 12 October 2017 'Compliance checklist for radio broadcast content' section 1.4 https://www.ofcom.org.uk/__data/assets/pdf_file/0023/44636/compliance_checklist_for_radio_broadcasters.pdf

194 See, for example, LSE Media Policy Project blog: https://blogs.lse.ac.uk/mediapolicyproject/tag/media-literacy/ [accessed 25 March 2019]

195 The House of Lords Select Committee on Communications in its report Growing up with the Internet 2nd Report of Session 2016–17 (HL Paper 130), 21 March 2017 also emphasised the importance of guidance on design.

196 Children's Charities' Coalition on Internet Safety (CHIS), Digital Manifesto (2015), para 62, available: http://www.chis.org.uk/2015/03/29/launch-of-digital-manifesto (accessed 2 April 2019)

197 See Naomi Creutzfeldt and Chris Gill, The Impact and Legitimacy of Ombudsman and ADR Schemes in the UK, The Foundation for Law, Justice and Society, 2014

198 See Josh Constine, 'Whatsapp has an encrypted child porn problem', Tech Crunch, 20 December 2018, available: https://techcrunch.com/2018/12/20/whatsapp-pornography/ [accessed 25 March 2019]

199 Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/38/35), 6 April 2018, paras 26-28

200 Dan Raywood, 'The Dark Side of the GDPR' Infosecurity Magazine, 3 April 2018, available: https://www.infosecurity-magazine.com/magazine-features/dark-side-gdpr/ [accessed 21 March 2019].

201 For a review of types of action see e.g. C. Hodges, 'Collective Redress: The Need for New Technologies' (2019) 42 Journal of Consumer Policy 59, available: https://link.springer.com/article/10.1007/s10603-018-9388-x (accessed 2 April 2019)

202 Competition and Markets Authority Guidance, What are super complaints?, May 2015, available: https://www.gov.uk/government/publications/what-are-super-complaints/what-are-super-complaints [accessed 25 March 2019]

203 For a discussion of Article 80 GDPR see e.g. F. Sorace, 'Collective Redress in the General Data Protection Regulation: An Opportunity to Improve Access to Justice in the European Union?', 7/2018 Working Papers Jean Monnet Chair (2018), available at http://diposit.ub.edu/dspace/handle/2445/123425

204 Law Commission, Abusive and Offensive Online Communications Scoping Report, November 2018, available: https://www.lawcom.gov.uk/abusive-and-offensive-online-communications/ [accessed 25 March 2019]

205 David Pierson and Tracey Lien, 'Facebook CEO Mark Zuckerberg shows support for the idea of regulation — but not the particulars', LA Times, April 2018 https://www.latimes.com/business/technology/la-fi-tn-zuckerberg-facebook-regulation-20180411-story.html [accessed 25 March 2019]

206 CNIL announcement, 21 January 2019: https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc [accessed 25 March 2019]

207 Legal Services Board, Regulatory sanctions and appeals processes: an assessment of the current arrangements, March 2014, available at: https://www.legalservicesboard.org.uk/projects/thematic_review/pdf/20140306_LSB_Assessment_Of_Current_Arrangements_For_Sanctions_And_Appeals.pdf [accessed 25 March 2019]

208 Information Notices – S142 Data Protection Act 2018

209 Prest v. Petrodel [2013] UKSC 34

210 Company Directors Disqualification Act 1986, available: https://www.legislation.gov.uk/ukpga/1986/46/contents [accessed 25 March 2019]

211 ICO News, 'SCL Elections prosecuted for failing to comply with enforcement notice', 9 January 2019, available: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/01/scl-elections-prosecuted-for-failing-to-comply-with-enforcement-notice/

212 The Privacy and Electronic Communications (EC Directive) Regulations 2003, available: http://www.legislation.gov.uk/uksi/2003/2426/contents/made [accessed 25 March 2019]

213 Fraud Act, 2006, available: https://www.legislation.gov.uk/ukpga/2006/35/contents [accessed 25 March 2019]

214 Bribery Act 2010, available: https://www.legislation.gov.uk/ukpga/2010/23/contents [accessed 25 March 2019]

215 HMRC, Tackling tax evasion: Government guidance for the corporate offences of failure to prevent the criminal facilitation of tax evasion, September 2017, available: https://www.gov.uk/government/publications/corporate-offences-for-failing-to-prevent-criminal-facilitation-of-tax-evasion [accessed 25 March 2019]

216 Criminal Finance Act, 2017, available: http://www.legislation.gov.uk/ukpga/2017/22/contents/enacted [accessed 25 March 2019]

217 Financial Conduct Authority, Senior Managers and Certification Regime, Updated February 2019, available: https://www.fca.org.uk/firms/senior-managers-certification-regime [accessed 25 March 2019]

218 HSE, Health and Safety at Work Statistics 2018, available: http://www.hse.gov.uk/statistics/enforcement.pdf (accessed 2 April 2019).
219 https://www.congress.gov/bill/115th-congress/house-bill/1865
220 Occupiers Liability Act 1957, available: https://www.legislation.gov.uk/ukpga/Eliz2/5-6/31/contents [accessed 25 March 2019]
221 Defective Premises Act 1972, available: http://www.legislation.gov.uk/ukpga/1972/35 [accessed 25 March 2019]
222 Committee on Standards in Public Life, The 7 principles of public life, May 1995, available: https://www.gov.uk/government/publications/the-7-principles-of-public-life [accessed 25 March 2019]
223 HMG, Communications White Paper: a new future for Communications, 2000, available: https://webarchive.nationalarchives.gov.uk/20100407191943/http:/www.culture.gov.uk/images/publications/communicationswhitepaper_fullreport.pdf [accessed 25 March 2019]
224 BBFC, Age Verification Regulator Guidance, 2019, available: https://www.ageverificationregulator.com/industry/guidance/ [accessed 25 March 2019]
225 HSE, RR610 - The nature, causes and consequences of harm in emotionally-demanding occupations, 2008, available: http://www.hse.gov.uk/research/rrhtm/rr610.htm [accessed 25 March 2019]
226 Science and Technology Committee (Commons), Impact of social media and screen-use on young people's health 31 January 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/82202.htm [accessed 25 March 2019]